

(1)



⑯ BUNDESREPUBLIK

DEUTSCHLAND



DEUTSCHES  
PATENT- UND  
MARKENAMT

⑯ Offenlegungsschrift

⑯ DE 199 14 808 A 1

⑯ Int. Cl.<sup>7</sup>:

C 07 H 21/00

C 07 H 1/00

C 07 B 61/00

C 12 N 15/63

C 12 Q 1/68

⑯ Aktenzeichen: 199 14 808.2

⑯ Anmeldetag: 31. 3. 1999

⑯ Offenlegungstag: 19. 10. 2000

⑯ Anmelder:

Rauhe, Hilmar, 45128 Essen, DE

⑯ Vertreter:

Reinert, P., Dipl.-Biol. Dr.rer.nat., Pat.-Anw., 50935  
Köln

⑯ Erfinder:

Rauhe, Hilmar, 45128 Essen, DE; Feldkamp, Udo,  
47057 Duisburg, DE; Banzhaf, Wolfgang, 44359  
Dortmund, DE; Howard, Jonathan, 50931 Köln, DE

⑯ Entgegenhaltungen:

US 58 43 661

Chem. Abstr. 125(1996)194609y;

**Die folgenden Angaben sind den vom Anmelder eingereichten Unterlagen entnommen**

Prüfungsantrag gem. § 44 PatG ist gestellt

⑯ Informationstragende Polymere

⑯ Die Erfindung betrifft Verfahren zur Herstellung informationstragender Polymere, nach diesen Verfahren erhältliche informationstragende Polymere; Verfahren zur Isolierung, zur Vervielfältigung und zum Auslesen solcher informationstragender Polymere; polymere Datenspeicher und DNA-Computer, die informationstragende Polymere umfassen sowie die Verwendung informationstragender Polymere zur Herstellung von Molekulargewichtsstandards, als Marker oder Signaturen, zur Verschlüsselung von Information, als molekulare Kleber oder zur Herstellung oder Bearbeitung kleinstster molekularer Strukturen.

DE 199 14 808 A 1

DE 199 14 808 A 1

## Beschreibung

Die Erfindung betrifft Verfahren zur Herstellung informationstragender Polymere, nach diesen Verfahren erhältliche informationstragende Polymere; Verfahren zur Isolierung, zur Vervielfältigung und zum Auslesen solcher informationstragender Polymere; polymere Datenspeicher und DNA-Computer, die informationstragende Polymere umfassen sowie die Verwendung informationstragender Polymere zur Herstellung von Molekulargewichtsstandards, als Marker oder Signaturen, zur Verschlüsselung von Information, als molekulare Kleber oder zur Herstellung oder Bearbeitung kleinsten molekularen Strukturen.

Es ist bekannt, Nukleinsäuren wie DNA (Deoxyribonucleic Acid = Desoxyribonukleinsäure) und RNA (Ribonucleic Acid = Ribonukleinsäure) auch außerhalb von Organismen zur Informationsverarbeitung zu verwenden. Es wurden bisher verschiedene Ansätze, DNA Moleküle zur Informationsverarbeitung zu verwenden, vorgestellt. Die WO 97/07 440 und [L. M. Adleman, Molecular Computation of Solutions to Combinatorial Problems, Science, 266, 1021–1024, (1994)] beschreiben ein Verfahren, einen Graphen-Algorithmus ("Hamiltonian-Pfad-Problem") als Entscheidungsproblem mit Hilfe von DNA Molekülen zu berechnen. Der Algorithmus wird implementiert, indem die Kanten und Knoten des Graphen als DNA Sequenzen repräsentiert werden. Die Berechnung des Algorithmus wird dadurch vorgenommen, daß durch die Hybridisierung komplementärer DNA Sequenzen eine Menge von Pfaden des Graphen erzeugt werden.

Aus dieser Menge von Pfaden werden dann mit Hilfe molekularbiologischer Verfahren alle Pfade entfernt, die eine falsche oder keine Lösung des Algorithmus sind. Wurden alle Schritte erfolgreich ausgeführt, so muß am Ende entweder mindestens ein DNA Strang übrigbleiben, der die richtige Lösung des Problems enthält, oder es bleibt kein DNA Strang übrig, was bedeutet, daß der Algorithmus keine Lösung hat.

Das beschriebene Verfahren setzt voraus, daß genügend viele Moleküle zur Verfügung stehen, damit jede mögliche Lösung der jeweiligen Probleminstanz statistisch mindestens einmal in der Ausgangsmenge von Pfaden vorkommt. Dies kann aufgrund der statistischen Natur der Hybridisierungsvorgänge jedoch nicht garantiert werden. Außerdem können im ersten Schritt auch zyklische Graphen entstehen, was von vornherein die Menge falscher Lösungen vergrößert.

Der in WO 97/07 440 beschriebene Ansatz zeigt, daß Symbolverarbeitung überhaupt *in vitro* mit DNA Molekülen vorgenommen werden kann. Jedoch ist der beschriebene Ansatz in der Praxis kaum verwendbar: Ein Problem ist, daß der Algorithmus nicht deterministisch, sondern nur stochastisch ist, das gefundene Ergebnis ist damit nur mit einer bestimmten Wahrscheinlichkeit korrekt. Ein weiteres Problem ist, daß die Implementierung des Algorithmus nicht effizient (Laufzeit-optimal) ist, dadurch wird die Berechnung langsam. Es wird dargelegt, daß das verwendete Verfahren zur Lösung NP-vollständiger Probleme (einer Klasse von Problemen, für die nur deterministische Lösungsalgorithmen mit exponentieller Laufzeit bekannt sind und von der vermutet wird, daß es keine effizienten deterministischen Lösungsalgorithmen gibt) besonders gut geeignet sei, weil der Algorithmus aufgrund der Parallelität der ausgeführten Operationen nur lineare Laufzeit benötige. Diese Argumentation ist jedoch insofern fehlerhaft, als die lineare Laufzeit durch eine exponentielle Anzahl von Molekülen kompensiert wird. Die exponentielle Problemgröße bleibt also unverändert bestehen.

Insgesamt ist das in der WO 97/07 440 beschriebene Verfahren zu sehr eingeschränkt, um für molekulare Informationsverarbeitung über den beschriebenen Algorithmus hinaus verwendbar zu sein: Das Hamiltonian-Pfad-Problem ist fest kodiert, andere Algorithmen können damit nicht berechnet werden, jede Probleminstanz muß neu kodiert werden. Programmierung ist nicht vorgesehen, alle Schritte des Algorithmus werden von Hand ausgeführt. Ein Input-/Output-System ist nicht vorgesehen. Insgesamt ist das Verfahren zur Programmierung von Algorithmen oder zur Implementierung eines Computers nicht geeignet. Die WO 97/29 117 und [Frank Guarnieri, Makiko Fliss, Carter Bancroft, Making DNA Add, Science, 273, 220–223, (1996)] beschreiben ein Verfahren, Additionen mit Hilfe von DNA Molekülen auszuführen. Die Addition erfolgt als Schiebeoperation mit Überlauf-Übertrag *in vitro* mit DNA Molekülen und Primer-Elongation.

Das beschriebene Verfahren beschreibt ausschließlich Additionen. Jedoch ist selbst die Verwendung als Addierer aus mindestens zwei Gründen ungünstig: Zum einen ist Addition mit dem beschriebenen Verfahren nicht effizient (Laufzeit-optimal) implementiert. Zum anderen ist das verwendete System formal unvollständig, weil die durch Addition erzeugten Ergebnisse ihrerseits nicht mehr als Zahlen für Berechnungen verwendet werden können. Über die Addition hinausgehende Konzepte fehlen. Insgesamt ist das Verfahren zur Programmierung von Algorithmen oder zur Implementierung eines Computers nicht geeignet.

In [Qi Ouyang, Peter D. Kaplan, Shumao Liu, Albert Libchaber, DNA Solution of the Maximal Clique Problem, Science, 278, 446–449, (1997)] wird ein Verfahren beschrieben, einen Algorithmus zur Lösung des "Max-Clique-Problem's" zu implementieren. Der Algorithmus stammt aus der Graphentheorie und fällt unter die NP-vollständigen Probleme. Wie in [L. M. Adleman, Molecular Computation of Solutions to Combinatorial Problems, Science, 266, 1021–1024, (1994)] ist das Verfahren nur in der Lage, einen fest kodierten Lösungsalgorithmus zu berechnen. Auch hier wird das Ergebnis nur mit einer bestimmten Wahrscheinlichkeit erhalten. Das von den Autoren beschriebene Verfahren enthält keine weiterführenden Konzepte (etwa in Hinsicht der Implementierung anderer Algorithmen). Insgesamt ist das Verfahren zur Programmierung von Algorithmen oder zur Implementierung eines Computers nicht geeignet.

In [Erik Winfree, Xiaoping Yang, Nadrian C. Seeman, Universal Computation via Selfassembly of DNA: Some Theory and Experiments, Proceedings of the 2nd DIMACS Meeting on DNA-Based Computers, Princeton University, June 20–12, (1996)] wird erörtert, reguläre Grammatiken in DNA durch Verknüpfung von Oligonukleotiden mit "sticky ends" zu implementieren. Die Ausführungen sind jedoch rein theoretischer Natur und scheitern an wesentlichen, bisher ungelösten Problemen. Insbesondere wird nicht dargelegt, wie die zur Implementierung von Grammatiken benötigten Sequenzen konstruiert werden müssen. Gerade dies ist aber das entscheidende Problem, ohne dessen Lösung Grammatiken nicht implementiert werden können. Der Grund dafür liegt darin, daß die Sequenzen, die die Variablen und Termini einer Grammatik repräsentieren, sowohl eindeutig, als auch untereinander hinreichend unähnlich sein müssen. Außerdem müssen sie bestimmte strukturelle, chemische und physikalische Eigenschaften erfüllen. Andernfalls sind Fehlhybridisierungen zwischen Sequenzen zwangsläufig, die bei der Polymerisierung zu ungewollten Kettenverlängerungen und Kettenabbrüchen führen und damit das Funktionieren des Verfahrens unmöglich machen. Dieses Problem verschärft

sich mit der Anzahl der benötigten Sequenzen exponentiell.

Das beschriebene Verfahren wird von den Autoren selbst wieder verworfen, bzw. insofern als unzureichend bezeichnet, als es keine sonderlich interessanten Berechnungen erlaube (Erik Winfree, Xiaoping Yang, Nadrian C. Seeman, Universal Computation via Self-assembly of DNA: Some Theory and Experiments, Proceedings of the 2nd DIMACS Meeting on DNA Based Computers, Princeton University, June 20–21, (1996), S. 8, 2. Absatz, 1. Zeile). Weiterführende Experimente der Autoren stützen sich entsprechend auch nicht weiter auf reguläre Grammatiken und lineare Polymere, sondern auf kontextsensitive Grammatiken zur Erzeugung von DNA Gitterstrukturen [Erik Winfree, Furong Liu, Lisa A. Wenzler & Nadrian C. Seeman, Design and self-assembly of two-dimensional DNA crystals, Nature, 394, 539–544, (1998)]. Der von den Autoren beschriebene Ansatz ist womöglich zur Erzeugung DNAbasierter Nanostrukturen (Herstellung von Katalysatoren usw.) geeignet. Zur Informationsverarbeitung eignet er sich hingegen nicht, hier sind die erzeugten Gitterstrukturen eher hinderlich.

Zwar sprechen die Autoren von der Möglichkeit, das mathematische Konzept der "Wang tiles" durch die erzeugten Gitterstrukturen zu realisieren und damit potentiell auch für Berechnungen zu verwenden, sie bleiben aber bei der bloßen Erwähnung einer solchen Möglichkeit, ohne zu beschreiben, wie dies im Sinne einer Informationsverarbeitung genutzt werden kann.

Es ist überhaupt zweifelhaft, ob der von den Autoren beschriebene Ansatz überhaupt zu einer Informationsverarbeitung geeignet ist: Die Gitterstrukturen verhindern die Vervielfältigung informationstragender Sequenzen (z. B. durch Klonierung oder PCR = Polymerase Chain Reaction = Polymerase Kettenreaktion) und machen es unmöglich, die erzeugten Strukturen als Daten zu verwenden, weil diese nicht mehr, etwa durch PCR, ausgelesen werden können. Entsprechend sieht der gewählte Ansatz in der jetzigen Form konzeptionell gar keine Möglichkeit vor, die erzeugten Strukturen im Sinne einer Informationsverarbeitung, etwa als Daten, zu nutzen.

Aus dem Stand der Technik ist bisher kein Verfahren bekannt, das es erlaubt, DNA Moleküle für die Implementierung effizienter Algorithmen zu verwenden. Es ist kein Verfahren bekannt, Berechnungen automatisch (etwa in Art eines molekularen Computers) durchzuführen. Es ist auch kein Verfahren bekannt, Programme in Form von regulären Grammatiken mit molekularen Verfahren so zu implementieren, daß damit

- a) unterschiedliche (im Idealfall beliebige) reguläre Grammatiken implementiert werden können
- b) die mit regulären Grammatiken erzeugten Wörter einer Sprache ausgelesen werden können
- c) die mit regulären Grammatiken erzeugten Wörter einer Sprache für technische Zwecke (z. B. Informationsverarbeitung, Polymerchemie) weiterverwendet werden können.

Der vorliegenden Erfindung liegt daher die Aufgabe zugrunde, ein Verfahren bereitzustellen, das es erlaubt, Informationsverarbeitung auf molekularer Basis mithilfe regulärer Grammatiken vorzunehmen, ohne dabei auf eine oder wenige Grammatiken eingeschränkt zu sein. Das Verfahren soll kompatibel zu herkömmlichen Computern sein und eine Informationsverarbeitung erlauben, die teilweise auf molekularer Basis und teilweise auf der Basis herkömmlicher Computer stattfindet. Das Verfahren soll programmierbar und weitgehend automatisierbar sein. Die innerhalb des Verfahrens erzeugten Polymere sollen lesbar und für weiterführende Anwendungen und Verfahren verwendbar sein.

Diese Aufgabe wird gelöst durch ein Verfahren zur Herstellung informationstragender Polymere, das umfaßt,

- I. eine reguläre Grammatik  $G = (\Sigma, V, R, S)$  mit einem endlichen Terminalalphabet  $\Sigma$ , einer endlichen Variablenmenge  $V$ , einer endlichen Regelmenge  $R$  und einem Startsymbol  $S$  zu definieren;
- II. das NFR-Verfahren (Niehaus-Feldkamp-Rauhe-Verfahren) zur Herstellung von Monomersequenzen (Oligo- oder Polymere);
- III. mit dem NFR-Verfahren eine in Schritt I definierte Grammatik zu implementieren, indem damit Monomersequenzen hergestellt werden, die die Regelmenge  $R$  einer Grammatik  $G$  eindeutig darstellen;
- IV. aus den in Schritt III hergestellten Monomersequenzen für jede Regel der Regelmenge  $R$  von  $G$  ein die Regel repräsentierendes Oligomer (Algomer) zusammenzusetzen (Algomer-Assemblierung);
- V. die in Schritt IV zusammengesetzten Oligomere (Algomere) zu informationstragenden Polymeren zu verknüpfen (Symbolpolymerisation).

Im Rahmen der vorliegenden Erfindung werden folgende Definitionen und Abkürzungen verwendet:

**Algomer:** Doppelsträngiges Oligomer, das eine Regel einer gegebenen Grammatik repräsentiert. Algomere können miteinander zu Logomeren verknüpft werden.

**Auslese-PCR:** PCR, die zum Auslesen der in Logomeren enthaltenen Information verwendet wird.

**Bitpolymerisation:** Prozeß des Verkettens von Algomeren zu Logomeren, wenn es nur zwei verschiedene Elongatoren gibt.

**Elongator:** Algomer mit zwei Überhangsequenzen; kann mit Terminator oder Elongator ligieren und führt zu einer Kettenverlängerung.

**Grammatik:** Formalismus, der Sprachen beschreibt. Er basiert auf der formalen Theorie von Sprachen [Chomsky, N., Three models for the description of language, JACM, 2 : 3, 113–124, (1956)], [Chomsky, N., On certain formal properties of grammars, Inf. and Control, 2 : 2, 137–167, (1959)], [Chomsky, N., Formal properties of grammars, Handbook of Math. Psych., 2, 323–418, (1963)].

Eine Grammatik  $G$  beschreibt eine Sprache  $L(G)$ , das Alphabet dieser Sprache und deren Syntax. Nach einer Grammatik  $G$  können alle Wörter dieser Sprache erzeugt werden.

Eine Grammatik  $G$  ist ein Quadrupel  $(\Sigma, V, R, S)$  mit Terminalalphabet  $\Sigma$ , Variablenmenge  $V$ , Startsymbol  $S$  und Regelmenge  $R$ .

**Logomer:** Polymer, das symbolische Informationen trägt und durch die Verkettung von Algomeren erzeugt wurde. Entsprechend besteht ein Logomer aus sich wiederholenden Einheiten von Algomeren. Ein Logomer repräsentiert ein Wort

einer Sprache L (G), die von der entsprechenden Grammatik G erzeugt wird.

Monomer: Einzelmolekül. Mehrere Monomere können zu längeren Ketten verknüpft werden und so Oligomere und Polymere bilden.

Monomere sind im Fall von Desoxyribonukleinsäure die Nukleotide (Adenin, Cytosin, Guanin, Thymin wie auch sog. Basenanaloga wie Hypoxanthin etc.).

Oligomer: Kurzkettiges Molekül aus sich wiederholenden Einheiten (Monomeren). Auch kurze doppelsträngige Moleküle werden als Oligomere bezeichnet.

PCR: Polymerase Chain Reaction = Polymerase Kettenreaktion; Verfahren zur exponentiellen Vervielfältigung von DNA.

10 Die PCR benötigt ein DNA Template, das vervielfältigt wird, und zwei Primer, die jeweils gegenläufig im Template ansetzen und als Startpunkte einer DNA Polymerisation fungieren. Durch iterative Wiederholung von Schmelzen-Hybridisierung-Polymerisations-Zyklen wird das DNA Template vervielfältigt.

Polymer: Langkettiges Molekül aus sich wiederholenden Einheiten (Monomeren).

Regel: auch: Produktionsregel, Ersetzungsregel oder Ableitungsregel.

15 Beschreibt die Ersetzung von Symbolen durch andere. Durch die wiederholte Anwendung von Regeln können Symbolketten erzeugt werden.

Sequenz: Informatik: Abfolge von Zeichen;

Chemie: Abfolge kovalent verbundener Monomere;

Molekularbiologie: Abfolge kovalent verbundener Nukleotide.

20 Komplementarität: Zwei Sequenzen sind dann komplementär, wenn sie miteinander hybridisieren können. Im Falle von DNA sind z. B. die Sequenzen 5'att3' und 5'aaat3' komplementär, die Sequenz 5'acgt3' ist zu sich selbst komplementär. Startsymbol: Variable innerhalb einer Grammatik, von der ausgehend, durch Anwendung der Regeln der Grammatik, eine Symbolkette erzeugt werden kann.

Symbolpolymerisation: Prozeß des Verkettens von Algomenen zu Logomenen.

25 Terminal: Symbol einer Grammatik. Ein Terminal kann nicht weiter ersetzt (substituiert) werden. Terminalen sind die "Buchstaben" der Worte einer Sprache L (G).

Terminator: Algomer mit einer Überhangsequenz. Kann nur mit Elongator ligieren. Führt zum Kettenabbruch bei einer Symbolpolymerisation.

30 Variable: Symbol einer Grammatik. Eine Variable kann gemäß der Regel einer Grammatik von Terminalen, Variablen oder Kombinationen von Terminalen und Variablen ersetzt werden.

#### Kurzbeschreibung der Abbildungen

**Abb. 1:** Algomere, wie in der Erläuterung zu Schritt IV des erfundungsgemäßen Verfahrens beschrieben. Die Algome 35 re besitzen jeweils eine doppelsträngige Kernsequenz, die ein Terminal einer gegebenen Grammatik darstellt und mindestens eine einzelsträngige Überhangsequenz, die eine Variable der gegebenen Grammatik darstellt, so daß jeweils ein Algomer genau eine Regel der gegebenen Grammatik repräsentiert. X und Y sind keine Variablen, sondern Überhangsequenzen, die z. B. für eine Klonierung benutzt werden können. Mit Überstrichen dargestellte Buchstaben kennzeichnen Komplementarität. Gemäß der Definition sind A0A und A1A Elongatoren, XsA und AeY Terminatoren. Die hier dargestellten Algome 40 re repräsentieren die im Folgenden beschriebenen Grammatik zur Erzeugung binärer Zufallszahlen.

**Abb. 2:** Symbolpolymerisation, wie in der Erläuterung zum erfundungsgemäßen Schritt V beschrieben. Algome werden durch Verkettung (im Falle von DNA: Hybridisierung und Ligation) zu Logomenen verknüpft. Das Logomer Xs01010101eY beinhaltet eine terminierte Bitfolge, die durch die Überhänge X und Y in einer definierten Orientierung vervielfältigt werden kann.

**Abb. 3:** Muster, das die Symbolpolymerisation für binäre Zufallszahlen nach Gelelektrophorese und Färbung zeigt (Spuren 1-3). Aufgrund der zufälligen Länge der binären Zufallszahlen ergibt sich ein regelmäßiges Leitermuster. Spur 4 zeigt einen 50 bp Molekulargewichtsstandard (Gibco BRL, Life Technologies, Katalog Nr. 10416-014).

**Abb. 4:** Klonierung eines durch Symbolpolymerisation erhaltenen Logomers in einem Vektor, wie zu Vervielfältigung und Isolierung von Logomenen im Folgenden beschrieben.

**Abb. 5:** Schematische Darstellung eines Bandenmusters, das durch Auslesen eines binären Logomers durch PCR und nachfolgende Gelelektrophorese erhalten wird (im Folgenden beschrieben). Im gezeigten Beispiel haben die Algome 55 re eine Länge von 30 bp (Überhangsequenzen nur  $\frac{1}{2}$  gerechnet). Bei bekannter Länge des Logomers reicht es, jeweils nur 0en oder nur 1en auszulesen. Beide Bits auszulesen dient der Kontrolle. Das Verfahren kann auch für mehrwertige Logome 60 re (mehr als 2 Elongatoren) verwendet werden.

**Abb. 6:** Bandenmuster einer Gelelektrophorese nach PCR zum Auslesen von drei verschiedenen Logomenen. Die Logome 65 re wurden nach der im Folgenden beschriebenen Grammatik für Zufallszahlen beliebiger Länge erhalten. Als Zufallszahlen von unten nach oben gelesen ist a = 262, b = 97, c = 329. M (Spur 5) ist ein 50 bp Molekulargewichtsstandard (Gibco BRL, Life Technologies, Katalog Nr. 10416-014).

**Abb. 7:** Schematische Darstellung des Auslesens von Logomenen durch Restriktionsverdau wie im Folgenden beschrieben. Die Elongatoren tragen Restriktionsschnittstellen, die derart asymmetrisch angeordnet sind, daß der Restriktionsverdau von Logomenen in einem eindeutigen Schnittmuster von Fragmenten resultiert. Das Schnittmuster entspricht Banden verschiedener Länge, und kann durch Gelelektrophorese sichtbar gemacht werden. R1 und R2 sind unterschiedliche Restriktionsenzyme, x und y bezeichnen die Länge der nach Restriktionsverdau erhaltenen Fragmente. Zweckmäßig ist z. B. ein Verhältnis x : y von 1 : 2.

**Abb. 8:** Bandenmuster, die durch Gelelektrophorese nach dem Auslesen von Logomenen durch Restriktionsverdau erhalten werden. Spur 7 und 8 enthalten Molekulargewichtsstandards (Spur 7: 50 pb Leiter, Gibco BRL, Life Technologies, Katalog Nr. 10416-014; Spur 8: 10 bp Leiter)

# DE 199 14 808 A 1

**Abb. 9:** Schematische Darstellung eines polymeren Datenspeichers auf der Basis von Algomeren und Symbolpolymerisation wie im Nachfolgenden beschrieben. Algomere können an Ankermoleküle (AM), die an einem festen Träger (C) gebunden ist, polymerisieren. Das Schreiben (W) erfolgt durch die wiederholte Abfolge (ReW) von Hybridisierungs-Ligations-Restriktionszyklen (Hyb, Lig, Res). Durch Denaturieren oder Restriktion (Den/Dig) können die erhaltenen Logomere abgetrennt und ausgelesen werden.

**Abb. 10:** Vereinfachte Darstellung der Bestandteile eines DNA Desktop Computers wie im Nachfolgenden beschrieben. Im Einzelnen sind:

A: Oligo-Synthesizer, B: Thermozykler, C: Reaktionskammern, D: Pipettierzubehör, E: Gel, F: Scanner, G: Steuerungscomputer

Es kennzeichnen:

1: Zugabe von Oligonukleotiden, 2: Zugabe synthetisierter Oligomere in Reaktionskammern des Thermozyklers, 3: Zugabe von Lösungen und Molekülen, die zur Algomer-Assemblierung, Symbolpolymerisation, zur Isolierung von Logomeren und zum Auslesen benötigt werden.

**Abb. 11:** Verschlüsselung von Logomeren wie im Folgenden beschrieben. Sind die Terminatoren und die darin primenden Primer unbekannt, so ist das jeweilige Logomer verschlüsselt und kann nicht ausgelesen werden (A). Ist dagegen ein in einem Terminator primender Primer verfügbar oder die Sequenz eines Terminators bekannt, so kann das jeweilige Logomer gelesen werden (B).

**Abb. 12:** Y-förmiges Molekül, das als Terminator bei der asymmetrischen Verschlüsselung von Logomeren verwendet werden kann. Elongatoren können an das mit 2 markierte Ende angeknüpft werden, weitere, z. B. Y-förmige, Moleküle an die mit 1 und 3 bezeichneten Enden. Werden mehrere Y-förmige Moleküle miteinander verknüpft, so erhält man baumartige Strukturen.

**Abb. 13:** Logomer mit Terminatoren s und e, die als baumartige Strukturen realisiert sind.

**Abb. 14:** Mit einem Vektor V verknüpftes Logomer L mit baumartigen Terminatoren. Die Terminatoren sind so konstruiert, daß jeweils nur ein Ast des Terminators mit dem Vektor verknüpft werden kann.

**Abb. 15:** Markierung von Nukleinsäuren mit Logomeren wie im Nachfolgenden beschrieben: Um gentechnisch hergestellte oder veränderte Produkte zu markieren, wird ein Logomer mithilfe rekombinativer Techniken in einen nicht-transkribierten Bereich, z. B. vor den Promotor (P) eines zu markierenden Genes (G), eingesetzt.

**Abb. 16:** Markierung von Dokumenten mit Logomeren wie im Nachfolgenden beschrieben. Spur 1 zeigt den verwendeten Molekulargewichtsstandard (50 bp, Gibco BRL, Life Technologies, Katalog Nr. 10416-014), Spur 2 (0-Bits) und Spur 3 (1-Bits) das Bandenmuster des Auslesens von Logomer Nr. 330 durch PCR aus wässriger Lösung, die Spuren 4 und 5 dasselbe wie die Spuren 2 und 3, wobei hierbei jedoch das Logomer Nr. 330 nicht aus wässriger Lösung, sondern in ca.  $10^9$  Molekülen/ $\mu$ l getrocknet auf Papier (3M Post-It, 1 Stunde getrocknet) als Papierschnipsel (ca. 1 mm<sup>2</sup>) in die Auslese-PCR einging.

**Abb. 17:** Herstellung von Molekulargewichtsstandards durch Auslese-PCR, die ein unäres Logomer als Template enthält, wie im Nachfolgenden beschrieben. 1 zeigt die Anordnung verschachtelter Primer in der Auslese-PCR, 2 das nach Gelektrophorese der PCR-Fragmente erhaltene Bandenmuster.

**Abb. 18:** Verkleben von Flächen mit Nukleinsäuren wie im Nachfolgenden beschrieben. C bezeichnet die zu verklebende Fläche, Log die Logomere, die zum Verkleben an die Flächen gebunden werden. 1 zeigt das Verhalten von Flächen mit nicht-komplementären, 2 das Verhalten von Flächen mit komplementären Logomeren.

**Abb. 19:** Verkleben von Flächen mit Logomeren, Antikörpern und Liganden wie im Nachfolgenden beschrieben: C<sub>0</sub> und C<sub>1</sub> bezeichnen die zu verklebenden Flächen, die aus gleichem oder unterschiedlichem Material bestehen können. Die zu verklebenden Flächen sind mit Logomeren (Log) versetzt. An die Logomere können Proteine, z. B. Antikörper vom Typ Ab A und Ab B binden, wobei Ab A und Ab B gleich oder verschieden sein können. Ab A und Ab B sind ihrerseits durch einen Liganden (Li) miteinander verbunden.

**Abb. 20:** Verkleben von Flächen mit Proteinen z. B. Antikörpern ohne Verwendung von Logomeren. C<sub>0</sub> und C<sub>1</sub> bezeichnen die zu verklebenden Flächen, die aus gleichem oder unterschiedlichem Material bestehen können. Die Proteine vom Typ Ab A und Ab B binden direkt an die zu verklebenden Flächen und sind ihrerseits durch einen Liganden (Li) miteinander verbunden.

## Erläuterung zu I

### Wahl von Grammatiken

Eine Grammatik ist ein Quadrupel  $G = (\Sigma, V, R, S)$  mit einem endlichen Terminalalphabet  $\Sigma$ , einer endlichen Variablenmenge  $V$ , einer endlichen Regelmenge  $R$  und einem Startsymbol  $S$ . Für eine Sprache  $L(G)$ , die durch  $G$  beschrieben wird, können alle Wörter aus  $L$  als Wörter über dem Terminalalphabet  $\Sigma$  gebildet werden, indem man das Startsymbol  $S$  entsprechend den Regeln aus  $R$  ableitet.

Die Definition einer Grammatik  $G$  nach Schritt I des erfundungsgemäßen Verfahrens ist insofern frei, als beliebige, endliche Mengen von Terminalen und Variablen kodiert werden können. Erfundungsgemäß bevorzugt ist jedoch die Definition von Grammatiken, die die Binärdarstellung von Daten gestattet, insbesondere von Grammatiken mit

$$\Sigma := \{0, 1, s_0, s_1, s_2, \dots, s_{n-1}, s_n, e_0, e_1, e_2, \dots, e_{m-1}, e_m\} \text{ mit } n, m \in \mathbb{N} \text{ und } n, m \geq 0.$$

Diese ermöglichen die Herstellung binärer Logomere.

Zur Veranschaulichung wird im Folgenden eine erfundungsgemäß definierte reguläre Grammatik zur Erzeugung von Zufallszahlen beliebiger Länge in Binärdarstellung gezeigt:  
Es sei eine Grammatik  $G = (Z, V, R, S)$  mit Terminalalphabet  $\Sigma := \{0, 1, s, e\}$ , Variablenmenge  $V := \{A\}$ , Startsymbol  $S$  und Regelmenge  $R$ :

# DE 199 14 808 A 1

{  
S := sA  
A → 0A  
A → 1A  
5 A → e  
},  
wobei  
s := Start  
e := Ende  
10 sind.

Mit dieser Grammatik können Wörter über dem Terminalalphabet

$$\Sigma := \{0, 1, s, e\}$$

15 gebildet werden. Alle Wörter der Sprache L beginnen mit "s" und hören mit "e" auf, dazwischen befindet sich eine zufällige Anzahl (auch 0) zufälliger Bits ("0" und "1"). Wörter aus der Sprache L, die nach R gebildet werden, sind beispielsweise:

- 20 a) S → sA → se  
b) S → sA → s1A → s1e  
c) S → sA → s0A → s00A → s001A → s0010A → s0010e  
d) S → sA → s1A → s10A → s100A → s1000A → s10000A → s100000e

25 Die als Wörter der Sprache L erzeugten Binärmuster können als beliebige jedoch eindeutige Darstellung von Datentypen, z. B. als Buchstaben, Zahlen, alphanumerische Zeichen oder Zeichenketten gelesen werden. Liest man sie als die Binärdarstellung von Zahlen, so sind die obigen Wörter in Dezimaldarstellung:

- 30 a) Ø (leer)  
b) 1  
c) 2  
d) 32.

## Erläuterung zu II

### 35 Das NFR-Verfahren zur Herstellung von Monomersequenzen

Das NFR-Verfahren (Niehaus-Feldkamp-Rauhe-Verfahren) ist ein Verfahren zur Herstellung von Monomersequenzen (Oligo- und Polymeren, z. B. Nukleinsäuren), das gewährleistet, daß die mithilfe des Verfahrens hergestellten Monomersequenzen:

- 40 a) eine eindeutige und einzigartige Abfolge von Monomeren haben;  
b) zueinander möglichst unähnlich sind und daher möglichst keine Fehlhybridisierungen untereinander eingehen;  
c) bestimmte strukturelle Eigenschaften aufweisen, z. B. ein bestimmtes Verhältnis der verschiedenen Monomere zueinander, das Enthaltensein oder Nicht-Enthaltensein bestimmter Teilsequenzen und eine maximale Übereinstimmung (Homologie) untereinander;  
45 d) bestimmte chemische Eigenschaften aufweisen, z. B. das Vorkommen bestimmter chemisch aktiver Teilsequenzen, die Einfluß auf chemische Reaktionen, im Falle von Nukleinsäuren insbesondere die Wechselwirkung mit Proteinen haben (Proteinbindungsstellen, Restriktionsschnittstellen, Stopcodons);  
50 e) bestimmte physikalische Eigenschaften aufweisen, z. B. eine bestimmte Schmelztemperatur.

Das NFR-Verfahren erlaubt die Herstellung eindeutiger, möglichst unähnlicher Monomersequenzen mit vordefinierten strukturellen, chemischen und physikalischen Eigenschaften. Es eignet sich zur Herstellung von Monomersequenzen für kontrollierte chemische Reaktionen, wie sie z. B. für eine molekularen Informationsverarbeitung benötigt werden. Es wird in Schritt III des erfundungsgemäßen Verfahrens für die Implementierung regulärer Grammatiken verwendet.

60 Das NFR-Verfahren ist eine erfundungsgemäße Fortbildung des Niehaus-Verfahrens zur Erzeugung eindeutiger, möglichst unähnlicher Sequenzen, das in [Jens Niehaus, DNA Computing: Bewertung und Simulation, Diplomarbeit am Fachbereich Informatik der Universität Dortmund, Lehrstuhl XI, 116–123, (1998)] beschrieben ist, und worauf hiermit in vollem Umfang Bezug genommen wird.

Das NFR-Verfahren ist insofern eine erfundungsgemäße Fortbildung des Niehaus-Verfahrens, als erst das NFR-Verfahren die Herstellung von Monomersequenzen erlaubt, wie sie für eine Herstellung *in vitro*, z. B. zur Implementierung von Grammatiken, benötigt werden. Die Gründe sind im Einzelnen:

- 65 a) Das Niehaus-Verfahren beschreibt nur die Konstruktion von Sequenzen aus Basissequenzen der Länge 6. Da die Länge von Basissequenzen die maximal zwischen verschiedenen Monomersequenzen möglichen Überlappungen beschreibt, daher unmittelbaren Einfluß auf das Hybridisierungsverhalten von Sequenzen und das Funktionieren der erfundungsgemäßen Schritte IV und V hat, muß das Verfahren jedoch für Basissequenzen beliebiger Länge verall-

# DE 199 14 808 A 1

gemeinert werden, was im NFR-Verfahren vorgenommen wurde.

- b) Das Niehaus-Verfahren erlaubt nur die Erzeugung von Sequenzen gleicher Länge, wohingegen das NFR-Verfahren Sequenzen unterschiedlicher Länge und unterschiedlicher Anzahl erzeugen kann. Letzteres ist für ein korrektes Hybridisierungsverhalten von Sequenzen und das Funktionieren der erfundungsgemäßen Schritte IV und V unerlässlich. 5
- c) Das Niehaus-Verfahren gewährleistet Eindeutigkeit und maximale Unähnlichkeit von Sequenzen nur für Sequenzen, die durch eine einmalige Anwendung des Verfahrens erzeugt wurden. Zwingend erforderlich ist jedoch, auch zu schon existierenden Sequenzen jeweils kompatible, d. h. eindeutige und maximal unähnliche Sequenzen erzeugen zu können. Das NFR-Verfahren kann im Gegensatz zum Niehaus-Verfahren zu beliebig vorgegebenen Sequenzen kompatibel (d. h. eindeutige und maximal unähnliche) neue Sequenzen erzeugen. Dabei kann das Verfahren beliebig oft auf dieselbe Menge von Sequenzen angewendet werden. 10
- d) Das Niehaus-Verfahren beschränkt zwar die maximale Länge gemeinsamer Teilsequenzen, jedoch nicht deren Anzahl. Daher können 2 beliebige nach dem Niehaus-Verfahren konstruierte Sequenzen mehrere gemeinsame Teilsequenzen beinhalten, was ungewollt zu einer hohen Homologie (Sequenzübereinstimmung) führen kann. Die Homologie hat ihrerseits unmittelbaren Einfluß auf das Hybridisierungsverhalten von Sequenzen und das Funktionieren der erfundungsgemäßen Schritte IV und V. Das NFR-Verfahren dagegen führt bei der Konstruktion von Sequenzen auch einen Homologievergleich durch und vermeidet dadurch ungewollte Fehlhybridisierungen. 15
- e) Das Niehaus-Verfahren erlaubt nicht die Integration bestimmter, vorgebbarer Sequenzen und Teilsequenzen. Solche Sequenzen sind jedoch für die Ausführung und Kontrolle bestimmter chemischer Reaktionen, z. B. kontrollierter enzymatischer Wechselwirkungen, zwingend notwendig. Beispiele für solche Sequenzen sind im Falle von Nukleinsäuren Proteinbindungsstellen, Restriktionsschnittstellen und Stopcodons. Dagegen erlaubt das NFR-Verfahren die Integration beliebiger Sequenzen und Teilsequenzen in die Herstellung von Monomersequenzen und gewährleistet gleichzeitig Eindeutigkeit und maximale Unähnlichkeit. 20
- f) Das Niehaus-Verfahren sieht keine Möglichkeit vor, Sequenzen mit bestimmten strukturellen, chemischen oder physikalischen Eigenschaften zu erzeugen. Dagegen enthält das NFR-Verfahren die Möglichkeit, Sequenzen mit bestimmten strukturellen, chemischen und physikalischen Eigenschaften herzustellen. Dazu zählen u. a. das Verhältnis verschiedener Monomere zueinander und die Schmelztemperatur. Auch dies ist für die Implementierung von Grammatiken *in vitro* eine unbedingte Voraussetzung. 25
- g) Das Niehaus-Verfahren erlaubt keine direkte Erzeugung regelrepräsentierender Sequenzen, wie sie zur Implementierung von Grammatiken benötigt werden. Dagegen ermöglicht das NFR-Verfahren die Erzeugung symbolrepräsentierender Sequenzen, die zu den für die Implementierung einer Grammatik benötigten regelrepräsentierenden Sequenzen verknüpft werden können. 30
- h) Nur mithilfe des NFR-Verfahrens können Sequenzen auch so konstruiert werden, daß die Eigenschaften der Eindeutigkeit, maximalen Unähnlichkeit sowie strukturelle, chemische und physikalische Eigenschaften auch für Teilsequenzen gelten. Z. B. lassen sich im Falle von Nukleinsäuren Monomersequenzen so herstellen, daß sowohl die Gesamtsequenz, als auch z. B. das erste Drittel der Sequenz einen 50%igen GC-Anteil haben. Diese Eigenschaft ist z. B. für das Funktionieren der Verkettung von Sequenzen, die pro Molekül für mehrere Hybridisierungereignisse vorgesehen sind (wie z. B. die im Folgenden beschriebenen Algomere), eine unbedingte Voraussetzung. Nur so lassen sich z. B. Sequenzen so konstruieren, daß unterschiedliche Hybridisierungereignisse pro Molekül gleichwahrscheinlich sind. 35
- i) Nur das NFR-Verfahren erlaubt die unmittelbare, automatische Herstellung von Monomersequenzen, z. B. durch einen Oligonukleotidsynthesizer. 40

Zur Herstellung von Monomersequenzen werden diese durch das NFR-Verfahren konstruiert und synthetisiert. Zur Herstellung von  $n$  Monomersequenzen wird das NFR-Verfahren, wie im Folgenden beschrieben, durchgeführt, wobei zur Beschreibung des Verfahrens folgende Abkürzungen benutzt werden: 45

$A$ : = Alphabet der Mächtigkeit  $a \in N$ ;

Im Falle von DNA ist  $A = \{a, c, g, t\}$  und  $a = 4$ .

$S_i$ : = Sequenz.

Folge von Elementen des Alphabets  $A$ , das einer Sequenz aus Monomeren entspricht.

$I_{seq}$ : = Länge der pro Verfahrenszyklus zu konstruierenden Sequenzen.

$S_{seq}$ : = Sequenz der Länge  $I_{seq}$ .

$S_{seq,k}$ : = ( $k-1$ )-tes Monomer der Sequenz  $S_{seq}$  ( $k \in N$   $k <= I_{seq}$ ).

$I_{bas}$ : = Länge der pro Verfahrenszyklus zur Konstruktion von Sequenzen verwendeten Basissequenzen.

Es gilt:  $0 < I_{bas} <= I_{seq}$ .

$S_{bas}$ : = Sequenz der Länge  $I_{bas}$  Basissequenz; Sequenzen der Länge  $I_{seq}$  werden aus Basissequenzen konstruiert.

$I_{ov}$ : = maximale Länge der Kette aufeinanderfolgender Monomere, die jeweils zwei durch das Verfahren erzeugte Sequenzen gemeinsam haben ("overlap").

$I_{bas} - 1$ .

$I_{ov}/I_{seq}$  = Verhältnis von maximal erlaubter Sequenzwiederholung zu Sequenzgesamtlänge;

1. Maß für die Fehlhybridisierungswahrscheinlichkeit.

$I_{bas}/I_{seq}$  = Verhältnis von Basissequenzlänge zu Sequenzgesamtlänge;

2. Maß für die Fehlhybridisierungswahrscheinlichkeit.

$M_{seq}$ : = Menge von Sequenzen.

$M_{bas}$ : = Menge aller Basissequenzen der Länge  $I_{bas}$ .

$M_{nobas}$ : = Teilmenge der Menge von Basissequenzen der Länge  $I_{bas}$ , die durch Dekomposition aus  $M_{seq}$  erhalten würd.

$|M_{nobas}|$ : = Mächtigkeit von  $M_{nobas}$  = Anzahl der Basissequenzen in  $M_{nobas}$ .

$n$ : = Anzahl pro Verfahrenszyklus zu konstruierender Sequenzen.

# DE 199 14 808 A 1

$n_{tot}$ : = Anzahl insgesamt in allen Verfahrenszyklen hergestellter Sequenzen.

$n_{max}$ : = maximale Anzahl pro Verfahrenszyklus konstruierbarer Sequenzen.

$h$ : = Homologie. Ein Maß für die Übereinstimmung zwischen zwei Sequenzen.

Es gilt:  $0 < h < 1$ .

- 5  $t_m$ : = Schmelztemperatur. Für eine DNA-Sequenz wird die Schmelztemperatur nach dem Nearest Neighbour Verfahren (siehe Breslauer, K. J., Frank, R., Blocker, H., Marky, L. A., Proc. Natl. Acad. Sci., 83, 3746–3750, 1989) bestimmt:

$$t_m = \Delta H / (\Delta S + R \times \ln(C/4)) - 273,15^\circ\text{C}$$

Dabei sind  $\Delta H$  und  $\Delta S$  Enthalpie und Entropie der DNA-Helix,  $R$  die molare Gaskonstante und  $C$  die Konzentration der DNA-Sequenz.

- 10 Das Herstellungsvorfahren erfolgt als ein Abfolge beliebig, jedoch endlich vieler Verfahrenszyklen, in denen  $n_{tot}$  Sequenzen konstruiert werden, und einer nachfolgenden Synthese, in der alle  $n_{tot}$  Sequenzen *in vitro* erzeugt werden.

Die maximale Anzahl  $n_{max}$  der pro Verfahrenszyklus konstruierbaren Sequenzen errechnet sich entsprechend den oben gegebenen Definitionen nach

15  $n_{max} = f(a, I_{bas}, I_{seq}) := \lfloor \frac{1}{2} \cdot ((a^{I_{bas}} - (a^{I_{bas}/2} - |M_{nobas}|)) / (I_{seq} - I_{ov}) \rfloor$ .

D. h. es lassen sich maximal  $n_{max}$  Sequenzen der Länge  $I_{seq}$  aus Elementen eines Alphabets  $A$  der Größe  $a$  (im Falle von DNA ist  $a = 4$ ) konstruieren, die in maximal  $I_{ov}$  zusammenhängenden Ketten von Monomeren übereinstimmen. Die Anzahl der tatsächlich pro Verfahrenszyklus erhaltenen Sequenzen kann geringer sein, wenn diese bestimmte physikalische, chemische oder strukturelle Eigenschaften aufweisen sollen, oder wenn zusätzlich zu bereits vorhandenen Sequenzen weitere Sequenzen hergestellt werden.

Im Einzelnen sind folgende Verfahrensschritte nötig:

- 25 1. Es werden  $n_{tot}$  Sequenzen in  $z$  Verfahrenszyklen konstruiert und in einer Menge  $M_{seq}$  gesammelt. Man beginnt mit der leeren Sequenzmenge  $M_{seq}$ .

2. Zur Sequenzmenge  $M_{seq}$  können einmalig, vor Beginn des Verfahrenszyklus, beliebige, jedoch eindeutige, Sequenzen hinzugefügt werden. Dies kann nützlich sein, um bestimmte Sequenzen hinzuzufügen, die aus Gründen der weiteren Anwendung in der Menge der hergestellten Monomersequenzen enthalten sein sollen. Es empfiehlt sich dabei, nicht zu viele und nicht zu lange Sequenzen hinzuzufügen, da das Verfahren für diese Sequenzen nicht dieselben Eigenschaften garantiert wie für die im Folgenden konstruierten Sequenzen.

- 30 3. Beginn des Verfahrenszyklus: Es werden die strukturellen, chemischen und physikalischen Eigenschaften festgelegt, die die herzustellenden Sequenzen erfüllen müssen. Strukturelle Eigenschaften sind zumindest das Enthaltensein oder Nicht-Enthaltensein bestimmter Teilsequenzen (ggfs. auch die Positionen, an denen die Teilsequenzen in den zu erzeugenden Sequenzen enthalten oder nicht enthalten sein sollen) und das Verhältnis der Anzahl der verschiedenen Monomere zueinander (im Falle von DNA: Verhältnis GC/AT). Chemische Eigenschaften sind mindestens das Vorkommen bestimmter Teilsequenzen, die Einfluß auf die chemische Wechselwirkung mit eigenen oder anderen Substanzen haben (im Falle von DNA z. B. bestimmte Proteinbindungsstellen, Restriktionsschnittstellen wie z. B. 5'gatac3' für EcoRV, die Stopcodons 5'tca3', 5'tta3', 5'cta3' usw.). Zu den festzulegenden physikalischen Eigenschaften zählt zumindest die Schmelztemperatur  $t_m$  der herzustellenden Sequenzen.

- 40 4. Es werden jetzt die Anzahl der pro Zyklus gewünschten Sequenzen  $n < n_{max}$ , nach obiger Formel  $I_{seq}$  der zu konstruierenden Sequenzen und  $I_{bas}$  der zur Konstruktion benötigten Basissequenzen festgelegt.  $I_{seq}$  und  $I_{bas}$  werden nach der oben angegebenen Formel so gewählt, daß die gewünschte Anzahl  $n$  zu konstruierender Sequenzen erreicht werden kann. Da  $I_{bas}/I_{seq}$  proportional zur Fehlhybridisierungswahrscheinlichkeit ist, werden  $I_{bas}$  und  $I_{seq}$  so gewählt, daß das Verhältnis  $I_{bas}/I_{seq}$  möglichst gering ist (Werte unter 0,3 sind empfehlenswert) und Werte für  $I_{seq}$  gute Hybridisierungsbedingungen garantieren. Für DNA sind temperaturabhängig beliebige Werte für  $I_{seq} > 0$  möglich.

- 45 5. Es wird die Menge  $M_{bas}$  aller Sequenzen des Alphabets  $A$  der Länge  $I_{bas}$  erzeugt. Die solcherart erzeugten Sequenzen werden als Basissequenzen bezeichnet. Es gibt dabei immer  $a^{I_{bas}}$  Basissequenzen. Diese Basissequenzen dienen der Konstruktion der herzustellenden Sequenzen. Jede Basissequenz bekommt einen Status "benutzt" oder "unbenutzt" zugewiesen. Es werden zunächst alle als unbenutzt markiert.

- 50 6. Selbst-komplementäre Basissequenzen aus  $M_{bas}$  werden nun als benutzt markiert. Dabei gibt es genau  $a^{I_{bas}/2}$  zu sich selbst komplementäre Basissequenzen.

- 55 7. Sollten bereits Sequenzen in der Menge  $M_{seq}$  vorhanden sein, so können entweder dazu kompatible neue Sequenzen konstruiert werden oder die Sequenzen einer Teilmenge von  $M_{seq}$  kompatibel verlängert werden. Dazu wird innerhalb eines Dekompositionsverfahrens aus den bereits vorhandenen Sequenzen aus  $M_{seq}$  eine Menge  $M_{nobas}$  aller Teilsequenzen der in Schritt 4 vorgegebenen Länge  $I_{bas}$  der Sequenzen aus  $M_{seq}$  gebildet:

Setze  $M_{nobas} = \{\}$ .

Jede Sequenz  $S_{seq}$  aus  $M_{seq}$  wird nun in  $(I_{seq} - I_{ov})$  Dekompositionsschritten zerlegt:

Beginne mit  $i = 0$  mit dem Dekompositionsschritt:

60 Solange  $i < (I_{seq} - I_{ov})$ :

Bilde als neue Sequenz  $S_{neu}$  aus  $S_{seq,i}$  als Sequenz der Länge  $I_{bas}$  bestehend aus den Monomeren  $S_{seq,i}$  bis  $S_{seq,i+I_{bas}}$ :  $S_{neu} := S_{seq,i}, \dots, S_{seq,i+I_{bas}}$ . Füge  $S_{neu}$  der Menge  $M_{nobas}$  hinzu.

Setze  $i := i + 1$ .

Die auf diese Weise erhaltene Menge  $M_{nobas}$  ist definitionsgemäß eine Teilmenge von  $M_{bas}$ .

- 65 8. Alle Basissequenzen der Menge  $M_{bas}$ , die auch in  $M_{nobas}$  vorkommen, werden als benutzt markiert, so daß für die Konstruktion von Sequenzen genau  $(a^{I_{bas}} - a^{I_{bas}/2} - |M_{nobas}|)$  unbenutzte Basissequenzen übrig bleiben, aus denen nun maximal

$$n_{\max} = \left\lfloor \frac{\left( \left( a^{l_{\text{bas}}} \right) - \left( a^{l_{\text{bas}}/2} \right) - |M_{\text{nobas}}| \right)}{2} \right\rfloor$$

$$\quad \quad \quad (l_{\text{seq}} - l_{\text{ov}})$$

5

verschiedene neue Sequenzen konstruiert werden können, die untereinander und zu den bereits vorhandenen Sequenzen keine Basissequenzen bzw. deren Komplemente gemeinsam haben.

10

9. Aus den Basissequenzen wird ein gerichteter Graph konstruiert, dessen Knoten bestimmte Basissequenzen repräsentieren: Der Graph enthält genau  $a^{l_{\text{bas}}}$  Knoten, von denen bereits  $(a^{l_{\text{bas}}/2} + |M_{\text{nobas}}|)$  Knoten als benutzt markiert sind. Jeder Knoten ist mit einer Basissequenz  $S_{b(i)}$  assoziiert, die nicht zu sich selbst komplementär ist (im weiteren wird es keine Unterscheidung von Knoten und assoziierter Basissequenz geben). Es existiert eine Kante von  $S_{b(i)}$  nach  $S_{b(k)}$ , wenn die  $I_{\text{ov}}$  letzten Buchstaben von  $S_{b(i)}$  den  $I_{\text{ov}}$  ersten Buchstaben von  $S_{b(k)}$  entsprechen, wenn also gilt:  $S_{b(i),2}, \dots, S_{b(i),l_{\text{bas}}} = S_{b(k),1}, \dots, S_{b(k),l_{\text{bas}}-1}$ . Der Knoten  $S_{b(k)}$  wird Nachfolgeknoten von  $S_{b(i)}$  genannt. Der Knoten, der mit dem Komplement der Basissequenz assoziiert ist, die durch den Knoten  $S_{b(i)}$  kodiert wird, wird Komplementknoten zu Knoten  $S_{b(i)}$  genannt. Sequenzen der Länge  $I_{\text{seq}}$  werden gefunden, indem ein Pfad mit  $(I_{\text{seq}} - I_{\text{ov}})$  Knoten gesucht wird. Jeder Knoten darf maximal einmal benutzt werden. Außerdem darf für jeden Knoten, der in einem der Pfade liegt, der Komplementknoten in keinem Pfad vorkommen. Der Anfangsknoten des Pfades trägt wie die Sequenz einen Status "benutzt" bzw. "unbenutzt".

15

10. Aus dem Graphen konstruiert man Sequenzen nach folgendem Verfahren:

20

Kennzeichne alle unbenutzten Knoten des Graphen als unbenutzte Anfangsknoten. Für jedes  $s$  zwischen 1 und  $n$ : Solange Knoten  $S_{b(k)}$  existiert, der noch nicht als benutzer Anfangsknoten markiert ist:

25

Wähle einen unbenutzten Knoten  $S_{b(k)}$  aus.

Markiere Knoten  $S_{b(k)}$  als benutzten Anfangsknoten; außerdem:

Markiere Sequenz  $S_{b(k)}$  und sein Komplement als benutzt.

Nun wird eine neue Sequenz  $S_{\text{neu}}$  konstruiert, die neues Element für  $M_{\text{seq}}$  ist oder eine bestehenden Sequenz  $S_{\text{sub}}$  aus  $M_{\text{seq}}$  verlängert:

30

Setze  $S_{\text{neu},0} := S_{b(k),1}$ .

Falls sie als neue Sequenz konstruiert wird, setze  $i := 0$ .

Falls sie als Verlängerung einer bereits bestehenden Sequenz  $S_{\text{sub}}$  konstruiert wird, setze  $i := I_{\text{sub}}$ .

Solange  $i < I_{\text{seq}} - (I_{\text{ov}} - 1)$  und  $i >= 0$  gilt:

35

Existiert kein unbenutzter Nachfolgeknoten  $S_{b(m)}$  von Knoten  $S_{b(k)}$ , so markiere Knoten  $S_{b(k)}$  und dessen Komplementknoten als unbenutzt und setze  $i := i - 1$ .

Sonst wähle per Zufall einen unbenutzten Nachfolgeknoten  $S_{b(m)}$  von Knoten  $S_{b(k)}$  aus und markiere  $b_m$  und dessen Komplementknoten als benutzt. Setze zusätzlich:  $i := i + 1$ ,  $k := m$ ,  $S_{\text{neu},i} := S_{b(m),1}$ .

40

Wenn  $i = I_{\text{seq}} - (I_{\text{ov}} - 1)$  ist, ist die Sequenz  $S_{\text{neu}}$  fertig: sie besteht aus den Buchstaben  $S_{\text{neu},0}$  bis  $S_{\text{neu},(I_{\text{seq}} - I_{\text{ov}})}$ .

11. Zu jeder der in Schritt 10 erhaltenen Sequenzen werden die jeweiligen strukturellen, chemischen und physikalischen Eigenschaften im Vergleich zu den in Schritt 3 festgelegten Bedingungen ermittelt. Genügt die jeweilige Sequenz nicht den vorgegebenen Bedingungen, so wird sie gelöscht und die von ihr benutzten Knoten und Komplementknoten wieder als unbenutzt markiert.

45

12. Wenn die Anzahl der konstruierten Sequenzen nicht ausreicht, kann der Verfahrenszyklus von Schritt 3 oder 4 an beliebig oft wiederholt werden. Insbesondere ist es möglich, die strukturellen, chemischen und physikalischen Kriterien und die Werte für  $n$ ,  $I_{\text{seq}}$ ,  $I_{\text{bas}}$  pro Verfahrenszyklus jeweils neu zu setzen. So ist es möglich, Sequenzen mit unterschiedlichen strukturellen, chemischen und physikalischen Eigenschaften, sowie unterschiedlicher Länge zu konstruieren, die trotzdem kompatibel, also eindeutig und maximal unähnlich sind. Ansonsten geht man zum nächsten Schritt, der in vitro Synthese.

50

13. Die konstruierten Sequenzen werden in vitro, im Falle von Nukleinsäuren vorzugsweise mittels eines Oligonukleotidsynthesizers (z. B. ABI 392, ABI 398, ABI 3948 von Perkin-Elmer Applied Biosystems), synthetisiert. Dazu werden die Sequenzdaten der Steuerungseinheit des Oligonukleotidsynthesizers übermittelt und die Sequenzen als einzelsträngige Nukleinsäuren hergestellt. Die Sequenzen können auch kommerziell bestellt werden. Zweckmäßig sind PAGE-gereinigte Oligonukleotide (z. B. von ARK Scientific GmbH Biosystems, 64293 Darmstadt, erhältlich).

55

### Erläuterung zu III

#### Implementierung regulärer Grammatiken mit dem NFR-Verfahren

Die Herstellung von Monomersequenzen zur Darstellung der Regelmenge von Grammatiken in Schritt III des erfundengemäßen Verfahrens erfolgt mit Hilfe des NFR-Verfahrens (Niehaus-Feldkamp-Rauhe-Verfahren) nach Schritt II des erfundengemäßen Verfahrens. Zur Implementierung von Grammatiken werden nach dem NFR-Verfahren für  $r$  Regeln einer Grammatik  $G$  genau  $2r$  Monomersequenzen hergestellt, so daß die Monomersequenzen für die  $s$  dargestellten Symbole (Terminale und Variablen) und die Regeln  $R$  der Grammatik  $G$  eindeutig und zueinander möglichst unähnlich sind, sowie die geforderten strukturellen, chemischen und physikalischen Eigenschaften aufweisen.

60

Die Monomersequenzen werden dazu nach dem INFR-Verfahren so hergestellt, daß sie wie im Folgenden beschrieben zu Algomen zusammengesetzt werden können. Es werden dabei jeweils 2 Monomersequenzen zu einem Algomer zusammengesetzt, das genau eine Regel  $R$  einer Grammatik  $G$  repräsentiert. Dabei enthalten beide Monomersequenzen je-

65

# DE 199 14 808 A 1

weils Sequenzen, die die nach der Regel R erforderlichen zusammengehörenden Symbole (Terminale oder Variablen) enthalten.

Der Entwurf der Oligomere nach Schritt III des erfindungsgemäßen Verfahrens ist abhängig von der chemischen Natur der eingesetzten Monomere. Im bevorzugten Fall der aus Nukleotiden aufgebauten Oligomere wird folgendermaßen vorgegangen:

Algomer sind doppelsträngig und haben einen 5'-Strang (Oberstrang) und einen 3'-Strang (Unterstrang). Oberstrang und Unterstrang können die Verknüpfung jeweils einer Terminal- und einer Variablensequenz sein.

Es seien X, Z beliebige Variablen, S eine Variable, die als Startsymbol fungiert, y, z beliebige Terminalssymbole. Dann wird für jede Regel der Form:

- 10 a)  $S := yX$  ein Algomer konstruiert, das eine eindeutige doppelsträngige Kernsequenz y enthält, an die am 3'-Ende eine eindeutige einzelsträngige Überhangsequenz X angehängt ist; ":" bedeutet in diesem Fall, daß S und yX identisch sind, d. h. yX als Startermolekül fungiert.  
15 b)  $X \rightarrow yZ$  ein Algomer konstruiert, das eine eindeutige doppelsträngige Kernsequenz y enthält, an die am 5'-Ende eine eindeutige einzelsträngige Überhangsequenz X und am 3'-Ende eine eindeutige einzelsträngige Überhangsequenz Z angehängt ist. X kann dabei identisch mit Z sein.  
c)  $X \rightarrow z$  ein Algomer konstruiert, das eine eindeutige doppelsträngige Kernsequenz z enthält, an die am 5'-Ende eine eindeutige einzelsträngige Überhangsequenz X angehängt ist.

20 Algomer der Form  $S := yX$  und  $X \rightarrow z$  heißen Terminatoren ("Start", "Ende"), da sie zum Kettenabbruch während der Polymerisation im erfindungsgemäßen Verknüpfungsschritt V führen, der im Folgenden beschrieben wird; Algomer der Form  $X \rightarrow yZ$  heißen Elongatoren, weil sie zu einer Kettenverlängerung während der Polymerisation im erfindungsgemäßen Verknüpfungsschritt V führen.

Vorzugsweise umfassen die in den Schritten II und III des erfindungsgemäßen Verfahrens hergestellten Monomersequenzen Nukleotide, insbesondere Ribonukleotide, besonders bevorzugt Desoxyribonukleotide. Die in Schritt II des erfindungsgemäßen Verfahrens konstruierten Monomersequenzen können vorteilhafterweise bestimmte Sequenzen, z. B. Stopcodons, Erkennungssequenzen für Nukleinsäuren spaltende Enzyme (Restriktionsnukleasen), Erkennungssequenzen für Nukleinsäure-bindende Proteine u. a. enthalten, wie sie beispielsweise in [J. Sambrook, E. F. Fritsch, T. Maniatis, Molecular Cloning, A Laboratory Manual, (1989)], [Rolf Knippers, Molekulare Genetik, Georg Thieme Verlag, (1997)] und [Benjamin Lewin, Genes V, Oxford University Press, (1994)] beschrieben werden.

## Erläuterung zu IV

### Algomer-Assemblierung

35 Das Assemblieren der in Schritt III synthetisierten Sequenzen erfolgt in Schritt IV des erfindungsgemäßen Verfahrens. Im Rahmen der vorliegenden Erfahrung sind zur Durchführung des Schrittes IV alle dem Fachmann bekannten Techniken, Oligomersequenzen zusammenzusetzen, einsetzbar. Im erfindungsgemäß bevorzugten Fall des Einsatzes von Nukleotidsequenzen ist es zweckmäßig folgendermaßen vorzugehen:

- 40 a) Die zu Elongatoren gehörenden einzelsträngigen Sequenzen werden phosphoryliert. Dazu sind verschiedene dem Fachmann bekannte Protokolle möglich, z. B. das Folgende:  
In einem 20 µl Ansatz werden 16 µl einer synthetisierten 100 µM Sequenz, 2 µl Ligationspuffer (z. B. 50 mM Tris-HCl pH 7.5, 10 mM MgCl<sub>2</sub>, 10 mM Dithiothreitol, 1 mM ATP, 25 µg/ml BSA Bovine Serum Albumin, New England Biolabs) und 2 µl PNK (Polynukleotidkinase, z. B. von New England Biolabs, Katalognr. #201 S oder #201 L) 1 Stunde bei 37°C inkubiert. Volumina, Inkubationszeit und -temperatur können in dem Fachmann bekannter Weise variiert werden.  
b) Jeweils in einem Ansatz werden die zu einem Algomer gehörenden Einzelstränge (Ober- und Unterstrang) zu einem fertigen Algomer hybridisiert. Für Elongatoren können einfach die Phosphorylierungsansätze von Ober- und Unterstrang verwendet werden. Für die Hybridisierung sind mehrere Protokolle möglich; bevorzugt ist ein Denaturierungsschritt bei etwa 95°C zu Anfang (dieser deaktiviert gleichzeitig die PNK in den Ansätzen der Elongatoren) und eine langsame Hybridisierung. Z. B. kann für Oligomere der Länge 30 folgendes Protokoll verwendet werden, das gemäß den Kenntnissen des Fachmanns variiert werden kann:  
In einem 40 µl Ansatz werden jeweils 20 µl des Oberstranges (100 µM) und 20 µl des Unterstranges (100 µM) in einem Thermozyklator 5 Minuten auf 95°C erhitzt, 5 Minuten auf 72°C inkubiert und dann 25 Minuten jeweils 1°C pro Minute abgekühlt.

## Erläuterung zu V

### Herstellung von Logomeren durch Verkettung von Algomeren

#### Symbolpolymerisation

65 In Schritt V des erfindungsgemäßen Verfahrens werden die Algomeren zu längerkettigen, informationstragenden Polymeren (Logomeren) verknüpft. Da Algomer die Regelmenge R einer Grammatik G darstellen, entsprechen alle dabei erzeugten Logomere Wörtern der Sprache L(G), die durch die Grammatik G beschrieben wird.

Der geregelte Prozeß der Verkettung von Algomeren zu Logomeren wird als "Symbolpolymerisation" bezeichnet, im

# DE 199 14 808 A 1

Falle, daß die Terminalsymbole Bits repräsentieren, das heißt, im Falle, daß gilt:

$$\Sigma = \{0, 1, s_0, s_1, s_2, \dots, s_{n-1}, s_n, e_0, e_1, e_2, \dots, e_{m-1}, e_m\} \text{ mit } n, m \in \mathbb{N} \text{ und } n, m > 0$$

wird der Prozeß als "Bitpolymerisation" bezeichnet.

Im Rahmen der vorliegenden Erfindung sind zur Durchführung des Schrittes V alle dem Fachmann bekannten Techniken, Oligomere zu Polymeren zu verknüpfen, einsetzbar. Im erfindungsgemäß bevorzugten Fall des Einsatzes von Oligonukleotiden ist es zweckmäßig, die zu verknüpfenden Algomere in Gegenwart von Ligase zu inkubieren. Beispielsweise kann nach folgendem Protokoll vorgegangen werden, das gemäß den Kenntnissen des Fachmanns variiert werden kann:

In einem 27 µl Ansatz werden 1 µl 50 µM "Start"-Algomer, 1 µl 50 µM "Ende"-Algomer und 2× jeweils 10 µl 40 µM Elongator-Algomer (aus Schritt IV des erfindungsgemäßen Verfahrens erhalten), 1,5 µl 10 mM rATP und 3,5 µl T4 DNA Ligase mit 400 NEB Units/µl (z. B. Katalognr. #202S und #202L NEB) zwischen 4°C und 25°C für 2 bis 24 Stunden inkubiert. Andere Reaktionsvolumina funktionieren analog. Inkubationszeit und -temperatur können in dem Fachmann bekannter Weise variiert werden.

Abb. 3 zeigt das Resultat einer solchen Symbolpolymerisation.

Es können dabei grundsätzlich beliebig viele Algomere, die keine Terminatoren sind, verknüpft werden. Der Polymerisationsvorgang kommt für jedes einzelne Logomer dann zu einem Stillstand, wenn das entsprechende Molekül an seinen Enden jeweils ein Algomer trägt, das als Terminatormolekül ("Start", "Ende") fungiert.

## Vereinzelung und Vervielfältigung von Logomeren durch Klonierung

Ein weiterer Gegenstand der vorliegenden Erfindung ist ein Verfahren zur Isolierung und Vervielfältigung von informationstragenden Polymeren, die nach einem der vorhergehenden Ansprüche erhalten würden, das dadurch gekennzeichnet ist, daß man in Schritt V erhaltene informationstragende Polymere in Klonierungsvektoren ligiert, kompetente Zellen mit diesen Vektoren transformiert und die erfolgreich transformierten Bakterien anhand von Selektionsmarkern selektioniert.

Die aus Schritt V des oben beschriebenen erfindungsgemäßen Verfahrens (Symbolpolymerisation) erhaltenen Logomere können zum Zwecke der Vereinzelung und Vervielfältigung kloniert werden.

Dazu wird der verwendete Klonierungsvektor mit Restriktionsenzymen aufgeschnitten und ein Logomer hineinligiert. Das Verfahren stellt sicher, daß nur fertig polymerisierte (mit Terminatoren versehene) Logomere mit dem Zielvektor ligiert werden können. Korrekt in den Zielvektor ligierte Logomere bilden ein wieder ringförmiges Molekül. Die Vereinzelung von Logomeren erfolgt durch die Transformation von Bakterien mit dem Molekülgemisch, das das Resultat der Ligation ist. Die Vektoren tragen Selektionsmarker (vorzugsweise Antibiotika-Resistenz z. B. Ampicillin-Resistenz), mit denen erfolgreich transformierte Bakterien ausgewählt werden können. Da jedes Bakterium nur ein Plasmid exprimiert, ist jedes erfolgreich transformierte Bakterium Träger genau eines Logomers. Solcherart klonierte Logomere können dann zur weiteren Verwendung einfach und in großen Mengen (auf Fest- oder in Flüssigmedium, Fermentern) vervielfältigt werden.

Die aus Schritt V des erfindungsgemäßen Verfahrens erhaltenen Logomere werden zum Zwecke der Vereinzelung und Vervielfältigung in einem beliebigen Klonierungsvektor (Plasmid) kloniert, der Restriktionsschnittstellen trägt, die zu den Sequenzüberhängen der Terminatoren der Logomere kompatibel sind (z. B. HindIII, BamHI Erkennungssequenzen im Klonierungsvektor pBluescript II KS +/-, Stratagene, Katalog Nr. #212207). Die Klonierung erfolgt nach üblichen Laborprotokollen, wie z. B. in [J. Sambrook, E. F. Fritsch, T. Maniatis, Molecular Cloning, A Laboratory Manual, (1989)] beschrieben. Z. B. kann folgendes Protokoll verwendet werden, das gemäß den Kenntnissen des Fachmanns variiert werden kann:

### a) Restriktion und Präparation des Klonierungsvektors

Zur Klonierung können beliebige Klonierungsvektoren verwendet werden, wie sie für molekularbiologische Verfahren üblich sind. (Geeignet ist z. B. das Plasmid pBluescript II KS +/-, Stratagene, Katalog Nr. #212207). Das Plasmid wird einem Restriktionsverdau mit zwei Restriktionsenzymen unterworfen, die Überhänge erzeugen, die zu den Terminatoren der Logomere kompatibel sind. Als Restriktionsenzyme können z. B. BamHI und HindIII (z. B. von NEB, New England Biolabs, Katalog Nr.: #136S und #104S) verwendet werden. Dazu kann ein übliches Restriktionsprotokoll verwendet werden, das gemäß den Kenntnissen des Fachmanns variiert werden kann, z. B. das Folgende:

In einem 50 µl Ansatz werden 30 µl Plasmid (0,7 µg/µl), 5 µl 10× Puffer (z. B. NEB2, New England Biolabs), 5 µl BamHI (100 units), 5 µl HindIII (100 units) und 5 µl BSA 10× für 1–2 Stunden bei 37°C inkubiert. Zur Kontrolle wird 1 µl des Ansatzes, sowie eine entsprechende Menge ungeschnittenes Plasmid auf einem 1% Agarose Gel (z. B. Ultra-Pure™ von Gibco BRL, Life Technologies, Katalog Nr.: 15510-027) elektrophoretisch aufgetrennt. Zur Präparation der DNA wird der Restriktionsansatz phenolisiert, die DNA gefällt und wiederaufgenommen:

- Ansatz auf 200 µl mit Aqua dest. auffüllen
- 200 µl Phenol zugeben, vortexen, bei 10000 g 3 Minuten zentrifugieren
- Überstand aufnehmen und 200 µl Chloroform zugeben, vortexen, bei 10000 g 3 Minuten zentrifugieren
- Überstand aufnehmen, mit 1/10 Vol. 3M NaAc ansäuern und mit 2,5× Volumen 100% EtOH versetzen, vortexen, für mindestens 15 Minuten auf -70°C
- mindestens 15 Minuten auf 10000 g zentrifugieren, Überstand abnehmen und verwerfen
- mit ca. 500 µl 70% EtOH waschen, für 5–10 Minuten auf 10000 g zentrifugieren, Überstand abnehmen und verwerfen

# DE 199 14 808 A 1

- Pellet trocknen und in Aqua dest. wiederaufnehmen, so daß das geschnittene Plasmid mit 100 ng/ $\mu$ l bis 1  $\mu$ g/ $\mu$ l vorliegt (höhere Konzentrationen sind auch möglich). Plasmid kann für weitere Ligationen nötigenfalls verdünnt werden.

5

## b) Ligation des Logomeres in den Klonierungsvektor

- Die in Schritt V des erfindungsgemäßen Verfahrens erhaltenen Logomere und die präparierten Klonierungsvektoren werden ligiert. Durch die Art und Weise der Präparation ist gewährleistet, daß möglichst nur die gewünschten Ligationen stattfinden: Die Terminatoren der Logomere tragen zwei verschiedene Überhangsequenzen, die zu den durch Restriktion entstandenen Überhangsequenzen des Klonierungsvektors kompatibel sind. Dadurch können die Logomere nur in einer definierten Richtung in den Klonierungsvektor ligieren. Außerdem sind die Terminatoren der Logomere nicht phosphoryliert, weshalb Logomere ausschließlich mit den Überhangsequenzen des Klonierungsvektors ligieren können. Die Ligation erfolgt nach einem üblichen Ligationsprotokoll wie z. B. in [J. Sambrook, E. F. Fritsch, T. Maniatis, Molecular Cloning, A Laboratory Manual, (1989)] beschrieben, das gemäß den Kenntnissen des Fachmanns variiert werden kann, beispielsweise wie folgt:

Das molare Verhältnis Logomere/Klonierungsvektor kann z. B. 500 betragen, in 10  $\mu$ l Gesamtvolume werden:

- 1  $\mu$ l Plasmid (10 ng/ $\mu$ l = 5 nM)  
20 0,5  $\mu$ l 4  $\mu$ M Logomere aus Schritt V des erfindungsgemäßen Verfahrens  
1  $\mu$ l 10x Ligationspuffer  
0,5  $\mu$ l T4 DNA Ligase (400 u/ $\mu$ l)  
7  $\mu$ l H<sub>2</sub>O

25 für 12 Stunden auf 16°C ligiert.

Das Protokoll kann gemäß den Kenntnissen des Fachmanns variiert werden. Z. B. ist es möglich, ein Protokoll zur TCL (Temperature Cycle Ligation, [Lund, A. H., Duch, M., Pedersen, F. S. Increased cloning efficiency by temperature-cycle ligation, Nucleic Acids Research, 24: (4), 800–801, (1996)]) zu verwenden.

Der erhaltene Ligationsansatz wird nun für die Transformation kompetenter Zellen verwendet, wie nachfolgend beispielhaft gezeigt:

## c) Transformation kompetenter Zellen

- 35 Die Transformation von Bakterien (kompetente Zellen, Hoststamm z. B. dH5 $\alpha$ , GIBCO BRL, Katalog Nr.: 18258-012) mit dem Ligationsansatz aus b) erfolgt nach einem der üblichen Protokolle, wie z. B. in [J. Sambrook, E. F. Fritsch, T. Maniatis, Molecular Cloning, A Laboratory Manual, (1989)] beschrieben wird und das gemäß den Kenntnissen des Fachmanns variiert werden kann, z. B. wie folgt:

- 40 – 200  $\mu$ l kompetenter Zellen ( $\sim 5 \times 10^7$  CFU = Colony Forming Units, bei -70°C gelagert) werden aufgetaut und auf Eis gestellt  
– Etwa 1 ng des Ligationsansatzes aus b) wird dazu pipettiert und 20–30 Minuten auf Eis stehen gelassen, alle 5 Minuten vorsichtig mixen  
– Hitzeschock: Ansatz 2 Minuten auf 42°C, zurück auf Eis  
– 0,8 ml LB-Medium (+ 0,02 M MgSO<sub>4</sub> + 0,01 M KCl) zugeben  
45 – Ansatz 20–60 Minuten im Reagenzglas auf 37°C rollen  
– 0,1–1 ml auf Agarplatte (mit Antibiotikum, z. B. Ampicillin) ausplattieren und über Nacht auf 37°C

Die Transformation fungiert dabei als Selektionsverfahren auf erfolgreich klonierte Logomere und als Isolationsverfahren einzelner Logomere, weil jede Bakterienzelle genau ein Plasmid exprimiert.

50 Außer dem genannten Verfahren der Klonierung, das erfindungsgemäß bevorzugt ist, können die in Schritt V des erfindungsgemäßen Verfahrens erhaltenen informationstragenden Polymere (Logomere) mit folgenden Verfahren isoliert und vervielfältigt werden:

## Isolation durch Hybridisierung und chromatographische Methoden

55

Hierbei wird das aus einer Symbolpolymerisation erhaltene Molekülgemisch über ein Trägermaterial gegeben, an das Oligomere ("Ankersequenzen") gebunden sind, die ihrerseits an die Terminatoren der Logomere binden können. Die gebundenen Logomere werden dann einzeln aufgenommen und können nach Bedarf vervielfältigt werden. Es sind hierzu verschiedene Verfahren einsetzbar:

60

a) Affinitätschromatographie; dabei werden Ankersequenzen, die zu den überhängenden Enden der Terminatoren kompatible Enden besitzen, z. B. an eine Hydroxylapatit-Säule gebunden. Die aus Schritt V des erfindungsgemäßen Verfahrens erhaltenen Logomere werden über die Säule gegeben, wodurch die Logomere durch Hybridisierung an die Ankersequenzen binden können. Die solcherart gebundenen Logomere werden dann einzeln aufgenommen.

65

b) Ankersequenzen werden an eine Membran (z. B. Gene Screen Plus, DuPont, Biotechnology Systems; Hybond-N, Amersham Life Sciences) kovalent gebunden. Die Membran ist gerastert, d. h. in einzelne Felder unterteilt. Pro Feld wird genau eine Ankersequenz kovalent an die Membran gebunden. Danach werden die aus Schritt V des erfindungsgemäßen Verfahrens erhaltenen Logomere über die Membran gegeben und mit den Ankersequenzen hybri-

# DE 199 14 808 A 1

disiert. Die Membran wird daraufhin in die einzelnen Felder des Rasters zerschnitten. Die einzelnen Felder, die nun maximal ein an die jeweilige Ankersequenz gebundenes Logomer enthalten, können jetzt in separate Gefäße (z. B. Eppendorf Tubes) überführt werden. Sodann können die Logomere durch Denaturieren von der Membran getrennt werden. Beim Denaturieren ist zu beachten, daß die Denaturierungstemperatur so gewählt ist, daß sie hoch genug ist, um Logomer und Ankersequenz zu trennen, jedoch nicht so hoch, daß das Logomer selbst vollständig aufschmilzt.

5

## Isolation durch Verdünnung

10

Bei der Isolation durch Verdünnung werden die Logomere, die als Gemisch aus einer Symbolpolymerisation erhalten werden, soweit verdünnt, daß in einem jeweiligen Zielvolumen statistisch genau ein Molekül enthalten ist. Dieses kann dann mit Hilfe der PCR aufgespürt und vervielfältigt werden. Die Verdünnung des Ausgangsvolumens in Zielvolumina kann dabei gleichzeitig geschehen, so daß ein Ausgangsvolumen auf n Zielvolumina verdünnt wird. Dazu wird erfindungsgemäß wie folgt vorgegangen:

15

Zu einem, aus Schritt V des erfindungsgemäßen Verfahrens erhaltenen, Gemisch von Logomeren in einem Ausgangsvolumen  $V_a$  wird ein Verdünnungsfaktor ermittelt, der in  $V_a$  enthaltene Logomere so auf n Zielvolumina  $V_z$  verdünnt, daß in jedem Zielvolumen  $V_z$  statistisch genau ein Logomer enthalten ist. Um den Verdünnungsfaktor zu ermitteln wird ein Aliquot des Ausgangsvolumens (z. B. 1 µl) in einer Verdünnungsreihe austitriert. Von jeder Verdünnung der Reihe wird dann ein Aliquot (z. B. 1 µl) als Template einer PCR-Reaktion verwendet, um zu ermitteln, in welchen Verdünnungen noch Logomere detektierbar sind. Erhält man solcherart die letzte Verdünnung, die noch Logomere enthält, und die Verdünnung, die schon keine mehr enthält, läßt sich ein ungefährer Verdünnungsfaktor angeben, der gerade noch ungefähr ein Logomer enthält. Gegebenenfalls kann dieser Verdünnungsfaktor durch Austitrieren der letzten Verdünnung, die noch Logomere enthält, genauer bestimmt werden. Dabei verwendet man entsprechend kleinere Verdünnungsschritte (verdünnt man z. B. in der ersten Verdünnungsreihe immer 1 : 10, so kann man in der 2. Verdünnungsreihe z. B. 1 : 2 verdünnen; das Verfahren der genaueren Bestimmung des Verdünnungsfaktors kann dabei prinzipiell mit beliebiger Genauigkeit angewendet werden).

20

Da die PCR der Verdünnungen dazu dient, festzustellen, ob sich überhaupt noch Logomere in der Verdünnung befinden, kann die PCR auf mindestens zweierlei Weise durchgeführt werden:

25

- a) Als Primer dienen zwei gegenläufige, in den Terminatoren primende Primer, z. B. der 5'-Strang des Start-Terminators und der 3'-Strang des Ende-Terminators. Ansonsten sind die PCR-Bedingungen wie weiter unten unter Vervielfältigung von Logomeren durch PCR beschrieben.
- b) Die PCR wird wie die zum Auslesen der Logomere durchgeführt, es reicht jedoch ein Ansatz mit einem Paar von Primern wie unter Auslesen von Logomeren mittels PCR im Folgenden beschrieben.

35

Zur Kontrolle werden die durch PCR erhaltenen DNA Fragmente durch Gelelektrophorese sichtbar gemacht. Dazu wird z. B. ein 2-4%iges Agarose-Gel (z. B. UltraPure™ von Gibco BRL, Life Technologies, Katalog Nr.: 15510-027) und als Molekulargewichtsstandard z. B. eine 50 bp-Leiter (Gibco BRL, Life Technologies, Katalog INr. 10416-014) verwendet. Das erhaltene Gel wird 5 Minuten in 0,001% Ethidiumbromid gefärbt und unter UV sichtbar gemacht.

40

## Vervielfältigung von Logomeren

45

Die aus einer Symbolpolymerisation erhaltenen Logomere können – je nach ihrer beabsichtigten Verwendung – vervielfältigt werden. Dies ist beispielsweise für die Visualisierung der gespeicherten Information oder die Weiterverwendung der Logomere als Marker zur Kennzeichnung von Stoffen und Gegenständen nötig.

## Vervielfältigung von Logomeren durch PCR

50

Bei diesem Verfahren werden Logomere durch FCR vervielfältigt. Das zu vervielfältigende Logomer dient als Template, die benötigten Primer primen jeweils gegenläufig in den Terminatoren (entweder 5'-Start und 3'-Ende, oder 5'-Ende und 3'-Start), so daß das Logomer vollständig vervielfältigt wird. Alternativ können die Primer, sofern das zu vervielfältigende Logomer von weiterer DNA umgeben ist, auch weiter außerhalb des Logomers primen.

55

Für die PCR-Ansätze, die in dem Fachmann bekannter Weise variiert werden können, werden z. B. folgende Reagenzien und Bedingungen verwendet:

dNTPs (z. B. Pharmacia Biotech, Katalog Nr.: 27-2035-01/2/3), Taq-Polymerase (z. B. Gibco-BRL, Katalog Nr.: 18038-034, 18038-042, 18038-067), PCR-Puffer, MgCl<sub>2</sub> (z. B. GIBCO-BRL, Katalog Nr.: 18067-017).

60

65

	Menge ( $\mu$ l)
dNTP, 10mM	1
5 PCR-Puffer 10x	5
MgCl <sub>2</sub> , 25mM	5
TAQ, 5u/ $\mu$ l	0,5
10 Primer 1, 10 $\mu$ M	1
Primer 2, 10 $\mu$ M	1
Template	1
15 (Logomer)	
H <sub>2</sub> O	35,5
Gesamtvolumen	50

20 Als PCR-Programm kann beispielsweise folgendes Protokoll (Primer: 30-mere) verwendet werden:

Schritt	Aktion	Temperatur	Dauer	Gehe zu	Anzahl
				Schritt	Wiederholungen
25 1	Denaturiere	95 °C	00:05:00		
2 2	Denaturiere	95 °C	00:00:30		
3 3	Annealing	68 °C	00:00:30		
30 4	Polymerisation	72 °C	00:00:30		
5 5	Gehe zu			2	29
6 6	Kühlen	4 °C	(beliebig)		
35 7	Ende				

Ein weiterer Gegenstand der Erfindung sind informationstragende Polymere, die gemäß den oben beschriebenen erfundungsgemäßen Verfahren erhältlich sind.

#### Auslesen der in Logomeren enthaltenen Information

40 Fertige Logomere können entweder mittels PCR (Polymerase Chain Reaction = Polymerasekettenreaktion), was erfundungsgemäß bevorzugt ist, oder mittels Restriktionsverdau gelesen werden. Dabei hat die PCR-Methode den Vorzug, bereits geringste Mengen DNA solcherart vervielfältigen zu können, daß die Logomere direkt nach der PCR und einer Gel-Auf trennung gelesen werden können. Im Falle von binären Logomeren kann das durch die PCR-Methode auf einem Gel erhaltene Bandenmuster unmittelbar als Binär code gelesen werden.

45 Ein weiterer Gegenstand der Erfindung ist daher ein Verfahren zum Auslesen von Information aus informationstragenden Polymeren, die wie oben beschrieben erhalten und/oder isoliert und vervielfältigt wurden, dadurch gekennzeichnet, daß man

- 50 a) n das informationstragende Polymer enthaltende Lösungen mit jeweils einem Paar gegenläufiger Primer versetzt, wobei n für die Zahl der als Elongatoren in dem Polymer enthaltenen Oligomere steht;
- b) mindestens n-1 PCR-Ansätze durchführt, wobei n für die Zahl der als Elongatoren in dem Polymer enthaltenen Oligomere steht und jeweils ein Primer eines jeden Paares in dem dem Elongator gegenüberliegenden Terminator primt und der andere Primer in dem Elongator selbst primt;
- 55 c) die aus der PCR erhaltenen Polymer-Fragmente nach ihrer Länge durch Elektrophorese auftrennt und d) das aus der Elektrophorese erhaltene Muster optisch ausliest.

Das Verfahren kann auch unter Verwendung verschiedenfarbiger fluoreszenzmarkierter Primer oder Nukleotide durchgeführt werden. Dadurch wird es möglich, mehrere Ansätze verschiedenfarbig zu markieren und mehrere Ansätze in einer Spur gelelektrophoretisch zu lesen. Überdies kann der Ausleseprozess dadurch mit modernen Sequenziermaschinen (z. B. von ABI erhältlich) automatisiert werden.

#### Auslesen von Logomeren mittels PCR

- 65 Um die in Logomeren enthaltenen Informationen sichtbar zu machen, können die Logomere mithilfe der PCR auslesen werden. Die Methode des Auslesens binärer Muster mit PCR wurde als DNA Typing bereits von [Jeffreys, Minisatellite repeat coding as a digital approach to DNA typing, Nature, 354, 204–209, (1991)] beschrieben. Die dort beschriebene Methode wird hier in einer abgewandelten Form als Auslese-PCR zum Auslesen der Logomere benutzt. Un-

# DE 199 14 808 A 1

terschiede zu der in [Jeffreys, Minisatellite repeat coding as a digital approach to DNA typing, Nature, 354, 204–209, (1991)] beschriebenen Methode bestehen darin, daß hier synthetisch erzeugte Templates verwendet werden, die nicht notwendig binäre Muster enthalten. Außerdem werden PCR- und Gel-Elektrophorese-Bedingungen so gewählt, daß das Ergebnis direkt vom Gel gelesen werden kann, ohne daß die als Bandenmuster erhaltenen Signale zusätzlich durch Blotting verstärkt werden müssen.

5

## a) PCR

Für das Auslesen eines Logomers werden für n Elongatoren der zugrundeliegenden Grammatik mindestens n-1, meistens aber n PCR-Ansätze gebraucht. Jeder PCR-Ansatz enthält das zu lesende Logomer als Template und außerdem ein Paar gegenläufiger Primer, von denen einer im Elongator, der andere im, dem Elongator gegenüberliegenden, Terminator primit (z. B. 5'-"Start" und 3'-0, siehe Abb. 5). Für die PCR kann jeder handelsübliche Thermozykler (z. B. PTC-100, MJ Research) verwendet werden. Als zweckmäßig für die Länge der Primer haben sich 20–30 bp erwiesen, jedoch sind auch andere Längen verwendbar. Je nach Primerlänge muß die Annealing-Temperatur entsprechend gewählt werden (kann z. B. mithilfe des Programmes Oligo 5.0 bestimmt werden). Zweckmäßige Annealingtemperaturen sind für 20-mere ca. 55°C, für 30-mere etwa 65°C bis 74°C. Für die PCR-Ansätze werden z. B. folgende Reagenzien und Protokolle verwendet, die in dem Fachmann bekannter Weise variiert werden können:  
dNTPs (Pharmacia Biotech, Katalog Nr.: 27-2035-01/213), Taq-Polymerase (Gibco-BRL, Katalog Nr.: 18038-034, 10  
18038-042, 18038-067), PCR-Puffer, MgCl<sub>2</sub> (GIBCO-BRL, Katalog Nr.: 18067-017)

10

15

20

	Menge ( $\mu$ l)	
dNTP, 10mM	1	
PCR-Puffer 10x	5	
MgCl <sub>2</sub> , 25mM	5	
TAQ, 5 $\mu$ l	0,5	
Primer 1, 10 $\mu$ M	1	
Primer 2, 10 $\mu$ M	1	
Template (Logomer)	1	
H <sub>2</sub> O	35,5	
Gesamtvolumen	50	

25

30

35

40

Um genügend DNA zu erhalten, so daß das erhaltene Bandenmuster nach der Gelelektrophorese direkt vom Gel gelesen werden kann, kann jeder PCR-Ansatz m mal angesetzt werden. Z. B. kann m = 4 sein.

Als PCR-Programm kann folgendes Protokoll (Primer: 30-mere) verwendet werden:

Schritt	Aktion	Temperatur	Dauer	Gehe zu		Anzahl	Schritt	Wiederholungen
1	Denaturiere	95 °C	00:05:00					
2	Denaturiere	95 °C	00:00:30					
3	Annealing	69,5 °C	00:00:30					
4	Polymerisation	72 °C	00:00:30					
5	Gehe zu			2		29		
6	Kühlen	4 °C	(beliebig)					
7	Ende							

45

50

55

Nach durchgeföhrter PCR ist es für die Lesbarkeit der Bandenmuster der nachfolgenden Gelelektrophorese evtl. zweckmäßig, möglichst viel amplifizierte DNA in möglichst geringem Volumen zu halten, das dann auf ein Gel geladen werden kann. Dazu können die Volumina auf geeignete Weise (z. B. mit einer SpeedVac, z. B. SpeedVac Concentrator, Savant) eingeengt werden.

60

## b) Gelelektrophorese

Die für jeden Elongator aus der PCR erhaltenen DNA Fragmente haben verschiedene Länge. Werden sie durch Elektrophorese aufgetrennt, so ergeben die unterschiedlichen Längenfragmente ein spezifisches Muster, anhand dessen sich das zu lesende Logomer identifizieren läßt (siehe Abb. 5 und Abb. 6). Für die Gelelektrophorese kann ein 4% Agarose-Gel (z. B. UltraPure™ von Gibco BRL, Life Technologies, Katalog Nr.: 15510-027), als Molekulargewichtsstandard z. B. eine 50 bp-Leiter (Gibco BRL, Life Technologies, Katalog Nr.: 10416-014) verwendet werden. Die Auf trennung

65

# DE 199 14 808 A 1

der DNA erfolgt in einer Gelkammer in geeigneter Weise, z. B. 1 : 45 Stunden bei 60 V. Hier sind die Parameter jedoch relativ frei wählbar, insbesondere kann die Gel-Laufzeit weiter verkürzt werden.

- Das erhaltene Gel wird 5 Minuten in 0,001% Ethidiumbromid gefärbt und unter UV sichtbar gemacht.
- Ein Beispiel für ein solcherart erhaltenes Gel findet sich exemplarisch unter Abb. 6.

5

## Auslesen von Logomeren durch Restriktionsverdau

Logomere können durch Restriktionsverdau ausgelesen werden: Dazu müssen die Elongatoren so konstruiert sein, daß sie asymmetrisch versetzte Restriktionsschnittstellen tragen. Jeder Elongator trägt dabei eine spezifische Restriktions-  
10 schnittstelle. Beispielsweise kann das Restriktionsenzym EcoRV (z. B. NEB, Katalog Nr. #195S) für 0-Elongatoren und SmaI (z. B. NEB, Katalog Nr. #141S) für 1-Elongatoren verwendet werden.

Zum Lesen wird das zu lesende Logomer in verschiedenen Restriktionsansätzen geschnitten. Dazu wird pro Elongator ein Restriktionsansatz mit dem jeweiligen Restriktionsenzym gebildet. Die aus der Restriktion erhaltenen DNA Frag-  
15 mente sind von verschiedener Länge. Werden sie durch Elektrophorese aufgetrennt, so ergeben die unterschiedlichen Längenfragmente ein spezifisches Muster, anhand dessen sich das zu lesende Logomer identifizieren läßt.

Das folgende Beispiel zeigt die DNA-Fragmentlängen aller binären Logomere mit 4 Bit ohne Terminatoren mit Elon-  
gatorlänge = 30 bp; Restriktionsschnittstelle pro Elongator nach 10 bp.

Binärzahl	0	1	
0000	10, 40, 40, 40, 30	160	
0001	10, 40, 40, 70	130, 30	
0010	10, 40, 80, 30	90, 70	*
0011	10, 40, 110	90, 40, 30	
0100	10, 80, 40, 30	50, 110	*
0101	10, 80, 70	50, 80, 30	
0110	10, 120, 30	50, 40, 70	
0111	10, 150	50, 40, 40, 30	
1000	50, 40, 40, 30	10, 150	
1001	50, 40, 70	10, 120, 30	
1010	50, 80, 30	10, 80, 70	
1011	50, 110	10, 80, 40, 30	*
1100	90, 40, 30	10, 40, 110	
1101	90, 70	10, 40, 80, 30	*
1110	130, 30	10, 40, 40, 70	
1111	160	10, 40, 40, 40, 30	

Im obigen Beispiel (4 Bits) erkennt man, daß die Längenmuster der Zahlen 0010 und 0100 bei Restriktion des 0-Bits nicht eindeutig ist. 0010 und 0100 lassen sich jedoch anhand der Längenmuster bei Restriktion des 1-Bits unterscheiden.

50 Im Einzelnen wurde zur Durchführung des obigen Beispiels folgendermaßen vorgegangen, wobei sich die Protokolle gemäß den Kenntnissen des Fachmanns variieren lassen:

### a) DNA Extraktion

Eine z. B. durch Vereinzelung und Vervielfältigung von Logomeren durch Klonierung erhaltene Bakterienkolonie wird zum Animpfen von 2–5 ml einer 37°C Übernachtkultur verwendet (z. B. LB-Medium mit 10 µg/l Ampicillin wie in [J. Sambrook, E. F. Fritsch, T. Maniatis, Molecular Cloning, A Laboratory Manual, (1989)] beschrieben). Als Klonierungsvektor können hier z. B. pGEM-T Easy (Promega, Katalog Nr. A1360) und binäre Logomere auch ohne Terminatoren verwendet werden. Aus der Übernachtkultur wird die das Logomer enthaltende Plasmid-DNA isoliert (z. B. mit Hilfe des Qiagen Plasmid Miniprep Kit, Qiagen, Katalog Nr. 12123, oder durch alkalische Lyse wie in [J. Sambrook, E. F. Fritsch, T. Maniatis, Molecular Cloning, A Laboratory Manual, (1989)] beschrieben).

### b) Restriktion

Die aus a) erhaltene Plasmid DNA wird für n Elongatoren in n Restriktionsansätzen geschnitten. Jeder Restriktions-  
65 ansatz enthält ein Restriktionsenzym, das in einem spezifischen Elongator schneidet und ein Restriktionsenzym, das das Logomer aus der Klonierungssite ausschneidet. Im verwendeten Beispiel wurden folgende Ansätze verwendet:

# DE 199 14 808 A 1

	Name	Menge	Volumen ( $\mu$ l)	
Plasmid	Logomer in pGEM-T	500ng/ $\mu$ l	6	
DNA	Easy (s.o.)			5
Puffer	NEB2	10x	1	
BSA		10x	1	
Enzym 1	Eco RV	10u/ $\mu$ l	1	
Enzym 2	Eco RI	10u/ $\mu$ l	1	10
Total			10	
				15
	Name	Menge	Volumen ( $\mu$ l)	
Plasmid	Logomer in pGEM-T	500ng/ $\mu$ l	6	
DNA	Easy (s.o.)			20
Puffer	NEB4	10x	1	
BSA		10x	0	
Enzym 1	Sma I	10u/ $\mu$ l	1	
Enzym 2	Eco RI	10u/ $\mu$ l	1	25
H <sub>2</sub> O			1	
Total			10	
				30

Die angegebenen Ansätze werden 1 Stunde auf 37°C inkubiert.

### c) Gelelektrophorese

Die aus b) erhaltenen DNA Fragmente werden elektrophoretisch aufgetrennt (z. B. 4% Agarose UltraPure™ von Gibco BRL, Life Technologies, Katalog Nr.: 15510-027), mit Ethidiumbromid (0,001%) gefärbt und unter UV sichtbar gemacht (siehe Abb. 8).

Die oben beschriebenen Verfahren ermöglichen beispielsweise die Implementierung binärer Logomere: Binäre Logomere werden mit Grammatiken erzeugt, in denen gilt:

$$\Sigma = \{0, 1, s_0, s_1, s_2, \dots, s_{n-1}, s_n, e_0, e_1, e_2, \dots, e_{m-1}, e_m\} \text{ mit } n, m \in \mathbb{N} \text{ und } n, m \geq 0.$$

Sie ermöglichen die einfachste und universellste Darstellung von Daten, wie sie auch von herkömmlichen Datenverarbeitungsmaschinen eingesetzt wird. Binäre Logomere sind Resultat einer Bitpolymerisation. Da abhängig von der jeweils gewählten Grammatik nahezu beliebige Symbole und Regeln kodiert werden können, können damit ebenso beliebige Daten und Datentypen erzeugt werden.

Die oben beschriebenen Verfahren ermöglichen beispielsweise die Implementierung von Zeichenalphabeten: Logomere können eingesetzt werden, um die Zeichen eines gewählten Alphabets zu kodieren. Die Zeichenlänge kann dabei unterschiedlich gewählt werden, sinnvoll ist es jedoch, die auch für herkömmliche Datenverarbeitungsanlagen benutzten Zeichenlängen (Halbbyte, 1-Byte, 2-Byte, 4-Byte, 8-Byte usw.) zu verwenden. Ebenso ist die Konvention für die Interpretation der dargestellten Zeichen wahlfrei. Die Zeichen können Zahlen, Buchstaben, alphanumerische Zeichen oder beliebige andere Datenstrukturen sein.

Die oben beschriebenen Verfahren ermöglichen beispielsweise die Implementierung eines 1-Byte Alphabets: Es sei eine Grammatik  $G = (\Sigma, V, R, S)$  mit Terminalalphabet  $\Sigma = \{0, 1, s, e\}$ , Variablenmenge

$$V: = \{S_0, S_1, S_2, S_3, S_4, S_5, S_6, S_7, S_8\}, \text{Startsymbol } S \text{ und Regelmenge } R: =$$

$$\begin{aligned} & \{ \\ & S := sS_0 \\ & S_0 \rightarrow 0S_1 \\ & S_0 \rightarrow 1S_1 \\ & S_1 \rightarrow 0S_2 \\ & S_1 \rightarrow 1S_2 \\ & S_2 \rightarrow 0S_3 \\ & S_2 \rightarrow 1S_3 \\ & S_3 \rightarrow 0S_4 \\ & S_3 \rightarrow 1S_4 \\ & S_4 \rightarrow 0S_5 \end{aligned}$$

$S_4 \rightarrow 1S_5$   
 $S_5 \rightarrow 0S_6$   
 $S_5 \rightarrow 1S_6$   
 $S_6 \rightarrow 0S_7$   
5     $S_6 \rightarrow 1S_7$   
 $S_7 \rightarrow 0S_8$   
 $S_7 \rightarrow 1S_8$   
 $S_8 \rightarrow e$   
} .

10

wobei:

 $s :=$  Start $e :=$  Ende Beispiel:

nach den Regeln der angegebenen Grammatik können alle 8-Bit Zeichen erzeugt werden:

15 z. B.: Erzeugung der 8-Bit 0, 00000000:

$S \rightarrow sS_0 \rightarrow s0S_1 \rightarrow s00S_2 \rightarrow s000S_3 \rightarrow s0000S_4 \rightarrow s00000S_5 \rightarrow s000000S_6 \rightarrow s0000000S_7 \rightarrow s00000000S_8 \rightarrow$   
 $s00000000e$

20 z. B.: Erzeugung der 8-Bit 255, 11111111:

$S \rightarrow sS_1 \rightarrow s1S_1 \rightarrow s11S_2 \rightarrow s111S_3 \rightarrow s1111S_4 \rightarrow s11111S_5 \rightarrow s111111S_6 \rightarrow s1111111S_7 \rightarrow s11111111S_8 \rightarrow s11111111e$

z. B.: Erzeugung der 8-Bit 85, 01010101:

25

$S \rightarrow sS_0 \rightarrow s0S_1 \rightarrow s01S_2 \rightarrow s010S_3 \rightarrow s0101S_4 \rightarrow s01010S_5 \rightarrow s010101S_6 \rightarrow s0101010S_7 \rightarrow s01010101S_8 \rightarrow$   
 $s01010101e$

Die für die Implementierung in vitro benötigten Sequenzen werden nach dem NFR-Verfahren, wie in den erfundungsgemäßen Schritten II und III beschrieben, hergestellt. Algomere und Logomere werden, wie in den erfundungsgemäßen Schritten IV–V beschrieben, hergestellt und die erhaltenen Logomere vorzugsweise mittels des unter Vereinzelung und Vervielfältigung von Logomeren durch Klonierung beschriebenen Verfahrens isoliert und vervielfältigt. Die solcherart erzeugten alphanumerischen Zeichen können jetzt für verschiedene technische Zwecke (z. B. die Markierung und Identifizierung von Stoffen) eingesetzt werden und, evtl. nach zusätzlichen technischen Schritten, gemäß dem erfundungsgemäßen Verfahren zum Auslesen von Information aus informationstragenden Polymeren (wie oben beschrieben) ausgelesen werden.

Da die hier erzeugten Binärmuster wahlfrei als Daten und Symbole gelesen werden können, können sie z. B. als alphanumerische Zeichen interpretiert werden. Werden sie als Zahlen interpretiert, so erzeugt die angegebene Grammatik 3-Bit Zufallszahlen. Aus einer genügend großen Menge zufälliger 8-Bit Binärmuster können alle 8-Bit Muster isoliert und als Bibliothek (z. B. zur Darstellung aller alphanumerischen Zeichen) angelegt werden.

Die oben beschriebenen Verfahren ermöglichen beispielsweise die Darstellung von Zeichenketten: Da Algomere so hergestellt werden können, daß sich unterschiedlich lange Logomere erzeugen lassen, ist es möglich, auch Zeichenketten darzustellen. Aus praktischen Gründen ist es jedoch zweckmäßig, nur wenige Zeichen pro Logomer vorzusehen. Z. B. kann ein einzelnes Logomer 4 Bit (Halbbyte) oder 8 Bit (1 Byte) enthalten. (Falls mehr Bit erforderlich sind, sollten zweckmäßigerweise immer Vielfache von 8 Bit = 1 Byte verwendet werden.) In Binärdarstellung können dann beliebige alphanumerische Zeichen in beliebiger Kodierung (z. B. ASCII, ANSI, ISO 8859-x, Unicode) dargestellt werden. Dabei kann sich ein Zeichen auch über mehrere Logomere erstrecken (Bsp.: Halbbyte-Darstellung, bei der sich jeweils ein 1-Byte Zeichen über zwei Logomere erstreckt, oder Unicode, bei der sich ein 16-Bit Zeichen über zwei 8-Bit oder vier 4-Bit Logomere erstreckt). Für die Darstellung von Zeichenketten ist es dann nötig, die einzelnen Zeichen mit Positionsinformation zu versehen. Dazu reicht es, in den jeweiligen Grammatiken verschiedene Terminatoren (z. B. verschiedene "Start"-Terminatoren) zu verwenden, deren Sequenzen jeweils die Positionsinformation repräsentieren. Beispiel: Es sei eine Grammatik  $G = (\Sigma, V, R, S)$  mit Terminalalphabet  $\Sigma = \{0, 1, e, s_0, s_1, s_2, \dots, s_{n-1}, s_n\}$ , Variablenmenge  $V = \{S_0, S_1, S_2, S_3, S_4, S_5, S_6, S_7, S_8\}$ , Startsymbol  $S$  und Regelmenge  $R =$

55

$\{$   
 $S := s_1S_0$   
 $S_0 \rightarrow 0S_1$   
 $S_0 \rightarrow 1S_1$   
 $S_1 \rightarrow 0S_2$   
60     $S_1 \rightarrow 1S_2$   
 $S_2 \rightarrow 0S_3$   
 $S_2 \rightarrow 1S_3$   
 $S_3 \rightarrow 0S_4$   
 $S_3 \rightarrow 1s_4$   
65     $S_4 \rightarrow 0S_5$   
 $S_4 \rightarrow 1S_5$   
 $S_5 \rightarrow 0S_6$   
 $S_5 \rightarrow 1S_6$

$S_6 \rightarrow 0S_7$   
 $S_6 \rightarrow 1S_7$   
 $S_7 \rightarrow 0S_8$   
 $S_7 \rightarrow 1S_8$   
 $S_8 \rightarrow e$   
}

5

wobei:

$s_i$  := Start, mit  $i = \{0, \dots, n\}$

$e$  := Ende.

10

Das mit dieser Grammatik erzeugte 1-Byte Alphabet reicht z. B. für alle alphanumerischen Zeichen des ASCII, ANSI oder ISO 8859-x Standards. Für eine Zeichenkette der Längen werden  $n$  verschiedene Terminatoren benötigt, so daß die mit dieser Grammatik erzeugten Zeichenketten wie folgt aufgebaut sind:

$s_0xe, s_1xe, \dots, s_{n-1}xe, s_nxe$

wobei  $x$  eine beliebige Binärdarstellung mit, in diesem Fall, 8 Bit ist.

15

Es bedeutet dann:

$s_0xe$  := Zeichen  $x$  an Position 0 (0.tes Zeichen)

$s_1xe$  := Zeichen  $x$  an Position 1 (1.tes Zeichen)

$s_2xe$  := Zeichen  $x$  an Position 2 (2.tes Zeichen)

usw.

20

Mit der Grammatik läßt sich beispielsweise die Zeichenkette "Elisabeth" im ASCII-Code darstellen:

$s_001000101e, s_101101100e, s_201101001e, s_301110011e, s_401100001e, s_501100010e, s_601100101e, s_701110100e, s_801101000e$

25

Alternativ kann auch  $\Sigma = \{0, 1, s, e_0, e_1, e_2, \dots, e_{n-1}, e_n\}$  gewählt werden, wodurch die Zeichenkette "Elisabeth" im ASCII-Code als:

$s_01000101e_0, s_01101100e_1, s_01101001e_2, s_01110011e_3, s_01100001e_4, s_01100010e_5, s_01100101e_6, s_01110100e_7, s_01101000e_8$

30

dargestellt wird. Eine weitere Möglichkeit ist  $\Sigma = \{0, 1, s_0, s_1, s_2, \dots, s_{n-1}, s_n, e_0, e_1, e_2, \dots, e_{n-1}, e_n\}$ , womit die Zeichenkette "Elisabeth" im ASCII-Code als:

$s_001000101e_0, s_101101100e_0, s_201101001e_2, s_301110011e_3, s_401100001e_4, s_501100010e_5, s_601100101e_6, s_701110100e_7, s_801101000e_8$

35

dargestellt wird.

In Halbbyte-Darstellung mit Positionsinformation würde die Zeichenkette "Elisabeth" im ASCII-Code dargestellt als:

40

$s_0100e, s_10101e, s_20110e, s_31100e, s_40110e, s_51001e, s_60111e, s_70011e, s_80110e, s_90001e, s_{10}0110e, s_{11}0010e, s_{12}0110e, s_{13}0101e, s_{14}0111e, s_{15}0100e, s_{16}0110e, s_{17}1000e$

45

Aufgrund der Positionsinformation, die durch die Sequenz der Terminatoren repräsentiert wird, können die einzelnen Zeichen völlig unabhängig voneinander verarbeitet werden und z. B. getrennt voneinander gelesen werden. Mit einem solchen Alphabet ist die Darstellung beliebiger Datentypen möglich. Z. B. können z. B. jeweils zwei Byte (0. + 1., 2. + 3., . . ., n. + n+1.) zu einem 2-Bytecode (z. B. Unicode) zusammengefaßt werden. Alphanumerische Zeichenketten können verwendet werden, um beliebige Bezeichner (Namen, Zahlen, Datum usw.) darzustellen.

50

Die für die Implementierung in vitro benötigten Sequenzen werden nach dem NFR-Verfahren, wie in den erfundungsgemäßen Schritten II und III beschrieben, hergestellt. Algomere und Logomere werden, wie in den erfundungsgemäßen Schriften IV-V beschrieben, hergestellt, die erhaltenen Logomere vorzugsweise mittels des unter Vereinzelung und Vervielfältigung von Logomeren durch Klonierung beschriebenen Verfahrens isoliert und vervielfältigt. Die solcherart erzeugten alphanumerischen Zeichen und Zeichenketten können jetzt für verschiedene technische Zwecke (z. B. die Markierung und Identifizierung von Stoffen) eingesetzt werden und, evtl. nach zusätzlichen technischen Schritten, gemäß dem erfundungsgemäßen Verfahren zum Auslesen von Information aus informationstragenden Polymeren (wie oben beschrieben) ausgelesen werden.

55

Die oben beschriebenen Verfahren ermöglichen beispielsweise die Implementierung eines Zufallszahlengenerators: Für bestimmte Probleme der Informatik (z. B. Simulationen) und der Mathematik werden Zufallszahlen benötigt. Diese werden im allgemeinen rechnergestützt durch Algorithmen erzeugt. Da diese Algorithmen jedoch deterministisch sind, handelt es sich bei den erzeugten Zufallszahlen nur um Pseudo-Zufallszahlen. Überdies sind die erzeugten Zahlenreihen aufgrund der unterschiedlichen Güte der Algorithmen unterschiedlich gut randomisiert. Für manche Anwendungen werden daher echte Zufallszahlen benötigt. Ist dies der Fall, muß ein physikalischer Prozeß in die Erzeugung von Zufallszahlen einbezogen werden. Ein solcher Prozeß ist etwa das Rauschen der Soundkarte in einem PC, das von entsprechender Software verarbeitet wird.

60

Mithilfe der hier beschriebenen Verfahren kann ein echter Zufallszahlengenerator implementiert werden, der Zufallszahlen sehr viel schneller erzeugt, als es mit herkömmlichen Verfahren möglich ist.

Der Zufallszahlengenerator wird mit folgender Grammatik implementiert (Grammatik für binäre Zufallszahlen beliebiger Länge):

65

G = ( $\Sigma$ , V, R, S) mit  $\Sigma$ : = {0, 1, s, e}, V: = {S},

R:

{

5 S: = sA

A  $\rightarrow$  0A

A  $\rightarrow$  1A

A  $\rightarrow$  e

}

10

wobei

s: = Start

e: = Ende

Mit dieser Grammatik können Wörter über dem oben angegebenen Terminalalphabet L = {0, 1, s, e} gebildet werden.

15 Alle Wörter der Sprache L beginnen mit "s" und hören mit "e" auf, dazwischen befindet sich eine zufällige Anzahl (auch 0) zufälliger Bits ("0" und "1").

Beispiel: Wörter aus der Sprache L, die nach R gebildet werden:

S  $\rightarrow$  sA  $\rightarrow$  se

20 S  $\rightarrow$  sA  $\rightarrow$  s1A  $\rightarrow$  s1e

S  $\rightarrow$  sA  $\rightarrow$  s0A  $\rightarrow$  s00A  $\rightarrow$  s001A  $\rightarrow$  s0010A  $\rightarrow$  s0010e

S  $\rightarrow$  sA  $\rightarrow$  s1A  $\rightarrow$  s10A  $\rightarrow$  s100A  $\rightarrow$  s1000A  $\rightarrow$  s10000A  $\rightarrow$  s100000e

25 Die als Wörter der Sprache L erzeugten Binärmuster können als Buchstaben, Zahlen, alphanumerische Zeichen oder Zeichenketten gelesen werden. Liest man sie als die Binärdarstellung von Zahlen, so sind diese Binärmuster in Dezimaldarstellung:

$\emptyset$  3 (leer)

1

30 2

32

und können als Zufallszahlen verwendet werden. Zur in vitro Implementierung eignen sich prinzipiell beliebige Sequenzen, solange sie eindeutig und zueinander maximal unähnlich sind. Z. B. können folgende Algomere verwendet werden:

35 sA:

agctttatatctccatggccctagtgaag

aatatacgtttaaacgggatcacttcaacc

40

A->0A:

ttggcgagatatacgcaccccttgctt

45

gctctatagttgcgtggggaaacgaaaaacc

A->1A:

50 ttggcgacccggaaaacaactattgttagt

gctggggcccttggataacgatcaaacc

A->e:

55

ttggtgccggagttggaaagcaactacgtg

acgcctcaaccccttcgttatgtcacctag

Die dazu benötigten Sequenzen sind handelsüblich. Sie können z. B. bestellt werden als 40 nmol, PAGE gereinigt (ARK-Scientific, Darmstadt) und z. B. 100  $\mu$ M in Aqua dest. aufgenommen werden (empfohlene Lagerung bei -20°C). Aus den Sequenzen werden, wie im erfundungsgemäßen Schritt IV beschrieben, Algomere hergestellt. Diese Algomere werden, wie im erfundungsgemäßen Schritt V beschrieben, zu Logomeren polymerisiert und können gemäß dem erfundungsgemäßen Verfahren zum Auslesen von Information aus informationstragenden Polymeren (wie oben beschrieben) ausgelesen werden. Derart erzeugte Zufallszahlen sind in Abb. 6 zu sehen.

65

Die oben beschriebenen Verfahren ermöglichen beispielsweise die Verschlüsselung von Logomeren:

Die in Logomeren enthaltene Information kann verschlüsselt werden. Dazu fungiert mindestens einer der für die Auslese-PCR benötigten Primer als geheimer Schlüssel (siehe Abb. 11). Bevorzugt ist dies einer der in den Terminatoren primentenden Primer. Ist die Sequenz dieses Primers (Schlüsselsequenz) und der Primer selbst nur autorisierten Zugreifern be-

kannt, so ist die in den derart verschlüsselten Logomeren enthaltene Information auch nur autorisierten Zugriffen zugänglich.

Um ein Auslesen der in den Logomeren enthaltenen Information ohne diesen Schlüssel zu verhindern, müssen weitere Vorkehrungen getroffen werden. Versuche, die Verschlüsselung zu brechen, können darauf beruhen, über die Vervielfältigung nicht-geheimer Sequenzen auch die Logomere zu lesen. Sofern bakterielle Vektoren zur Klonierung von Logomeren verwendet werden, könnte ein potentieller Angreifer z. B. versuchen, die gesamte Klonierungssite mit PCR zu vervielfältigen und dann mit Sequenzierung zu lesen oder über die zur Selektion benötigten Resistenzgene in die Klonierungssite "hineinzulesen". Auch könnten die Elongatoren selbst als Ansatzpunkte für ein PCR-basiertes Lesen der geheimen Sequenz(en) dienen.

Solchen Angriffsversuchen ist gemein, daß sie über nicht-geheime Sequenzen versuchen könnten, die Schlüsselsequenz zu entschlüsseln und darüber die verschlüsselte Sequenz zu lesen. Eine Abwehr derartiger Angriffsversuche kann dadurch erfolgen, daß zum Informationstransport über Logomere immer nur vollständig geheime Sequenzen verwendet werden. Jedoch ist dies evtl. zu aufwendig und kostenträchtig. Stattdessen oder ergänzend dazu können die informationstragenden Logomere mit Sequenzen (Scheinsequenzen) versetzt werden, die dieselben potentiellen Angriffspunkte (z. B. Bits, Resistenzgene) enthalten, jedoch jeweils andere Schlüsselsequenzen. Durch diese Scheinsequenzen laufen potentielle Angriffsversuche ins Leere, weil für den nicht-autorisierten Zugreifer alle Einzelsequenzen ununterscheidbar und dadurch die verschlüsselten Informationen verborgen sind. Mit je mehr Scheinsequenzen ein Logomer versetzt ist, um so schwieriger wird es, die Schlüsselsequenz zu brechen.

Das Verfahren entspricht einer molekularen Steganografie und wird exemplarisch anhand der Verschlüsselung von Zeichen eines 1-Byte Alphabets gezeigt:

Es sei die verwendete Grammatik  $G = (\Sigma, V, R, S)$  mit Terminalalphabet  $\Sigma := \{0, 1, s_{key}, s_0, s_1, s_2, \dots, s_{f-1}, s_f, e\}$ , wobei  $f \in \mathbb{N}$ ,  $f \geq 1$ .  $s_{key}$  ist der geheime Schlüssel. Für eine Schlüssellänge von  $I$  ist dann die maximale theoretische Verschlüsselung  $Key_{max} = a^I$ , die maximal verwendbare Verschlüsselung  $Key_{eff} = a^{1-d}$ , wobei  $d$  die Anzahl der Basen angibt, in der sich ein Scheinschlüssel von dem richtigen Schlüssel unterscheiden muß, damit in der Auslese-PCR kein Scheinschlüssel die richtige Sequenz lesen und umgekehrt der richtige Schlüssel kein Scheinlogomer (sondern nur das Ziellogomer) auslesen kann. Für hochspezifische PCR Konditionen kann  $d \geq 1$  werden. Weiterhin gibt  $Key_{imp} = \text{Anzahl Scheinschlüssel} + 1$  die tatsächlich verwendete Verschlüsselung und  $Key_{min} = 1 + x$  die minimale Verschlüsselung an. Dabei ist  $x$  die zu einer minimalen Verschlüsselung nötige Anzahl von Scheinlogomeren. Diese kann im Falle der Verwendung von  $n$  Logomeren als Informationsträger größer als  $n$  gewählt werden.

Für Zeichenketten ergibt sich automatisch eine zusätzliche Verschlüsselung, weil hier die für einen potentiellen Angreifer unbekannte Reihenfolge der Zeichen zusätzlich verschlüsselnd wirkt.

Beispiel: Für eine Grammatik mit Zeichenketten von  $n$  1-Byte Zeichen wird folgende Grammatik  $G$  verwendet:  
Es sei  $G = (\Sigma, V, R, S)$  mit Terminalalphabet  $\Sigma := \{0, 1, s_{key0}, s_{key1}, s_{key2}, \dots, s_{keyn-1}, s_{keyn}, s_0, s_1, s_2, \dots, s_{f-1}, s_f, e\}$ , wobei  $n, f \in \mathbb{N}$ ,  $n \geq 1$ .  $n$  gibt dabei die Anzahl der verwendeten echten Schlüssel,  $f$  die Anzahl der verwendeten Scheinschlüssel an; die verwendete Variablenmenge  $V$  und die verwendete Regelmenge  $R$  sind dabei wahlfrei und abhängig von der zu kodierenden Information.

Grammatiken mit anderer Zeichenkodierung funktionieren analog. Die für die Implementierung in vitro benötigten Sequenzen werden nach dem NFR-Verfahren, wie in den erfundungsgemäßen Schritten II und III beschrieben, hergestellt. Algomere und Logomere werden, wie in den erfundungsgemäßen Schritten IV-V beschrieben, hergestellt, die erhaltenen Logomere vorzugsweise mittels des unter Vereinzelung und Vervielfältigung von Logomeren durch Klonierung beschriebenen Verfahrens isoliert und vervielfältigt. Die solcherart erzeugten Logomere können, wie bereits als Auslese-PCR beschrieben, ausgelesen werden, wobei – wie beschrieben – einer der zum Auslesen benötigten Primer bevorzugt einer der in den Terminatoren primenden Primer, als geheimer Schlüssel fungiert. Mit der oben beschriebenen Grammatik werden zusätzlich zu den informationstragenden Logomeren "Scheinsequenzen" ("Scheinlogomere") erzeugt, die einzelnen nicht-autorisierten Auslese-Zugriff auf die in den Logomeren enthaltenen Informationen verhindern sollen. Das Verfahren kann für alle Anwendungen von Logomeren benutzt werden und hat u. a. den Vorzug, durch die Spezifität des als geheimer Schlüssel benutzten Primers sehr effektiv zu sein.

Basierend auf der beschriebenen Methode können symmetrische und asymmetrische Verschlüsselungsverfahren implementiert werden. Für ein symmetrisches Verschlüsselungsverfahren reicht es, daß sich die Teilnehmer A und B einer Kommunikation auf eine geheimgehaltene Schlüsselsequenz einigen. Teilnehmer B verschlüsselt eine als Logomer gespeicherte Nachricht an A, indem B Logomere nur mit den geheimen Terminatoren herstellt und seine Nachricht mit zusätzlichen Logomeren versetzt, die andere Terminatoren tragen. Nur A ist es dann möglich, das zum Auslesen benötigte Primerpaar bereitzustellen, da nur A die benötigte Terminatorsequenz kennt.

Ein asymmetrisches Verschlüsselungsverfahren erfordert dagegen zusätzlichen Aufwand, z. B. den Einsatz von Molekülen mit irregulären Formen (siehe z. B. Abb. 12 und Abb. 13): Will ein Kommunikationsteilnehmer B eine verschlüsselte Nachricht an A schicken, so bekommt er von A einen öffentlichen Schlüssel, den er zur Verschlüsselung der als Logomer repräsentierten Nachricht verwenden kann. Der öffentliche Schlüssel von A besteht aus einem Paar von Terminatoren und einem Pool zusätzlicher Scheinterminatoren mit ähnlichen Eigenschaften, jedoch abweichenden Sequenzen und anderen 3'-Überhangsequenzen. Von den echten Terminatoren ist lediglich die 3' Überhangsequenz bekannt, mit der sie an die – öffentlich bekannten – Elongatoren verknüpft werden kann. Zur Verschlüsselung einer Nachricht erzeugt B Logomere, wobei er in der Symbolpolymerisationsreaktion den von A öffentlich bereitgestellten Schlüssel (enthaltend Terminatoren und Scheinterminatoren) verwendet. Die Art und Weise der Terminatoren gewährleistet nun, daß lediglich A in der Lage ist, die solcherart verschlüsselte Nachricht wieder zu lesen: Da nur A die tatsächliche Sequenz der echten Terminatoren kennt, ist nur A in der Lage, die als Logomere verschlüsselte Nachricht über Auslese-PCR wieder zu lesen.

Weil der öffentliche Schlüssel potentiell über die bekannte Überhangsequenz zu den Elongatoren angreifbar ist (weil die Scheinterminatoren diese Überhangsequenz nicht besitzen dürfen, kann ein potentieller Angreifer eine beliebige Sequenz an die echten Terminatoren knüpfen, wodurch die Sequenz der Terminatoren identifiziert werden kann), werden als Terminatoren Moleküle auf der Basis irregulärer Formen, z. B. auf der Basis der in Abb. 12 gezeigten Y-förmigen

# DE 199 14 808 A 1

Moleküle verwendet. Die gezeigten Y-förmigen Moleküle lassen sich zu weiter verzweigten baumartigen Molekülen zusammensetzen. Die Wurzel eines solchen als Baum realisierten Terminatormoleküls enthält dann die zu den Elongatoren kompatible Überhangsequenz, während nur eine der Äste des Baumes die echte Terminatorsequenz enthält. Da die verschlüsselte Nachricht wie oben beschrieben zusätzlich mit Scheinlogomeren versetzt ist, kann dann nur A, der die echte Terminatorsequenz kennt, aus der Menge der Moleküle die richtige Nachricht durch Verknüpfen mit dem kompatiblen Vektor herausfiltern, wohingegen ein potentieller Angreifer ohne Kenntnis der echten Terminatorsequenz dies nicht kann.

Die oben beschriebenen Verfahren ermöglichen beispielsweise die Verschlüsselung mit Echtzeit-Identifizierung: Ohne das Auslesen der in verschlüsselten Logomeren enthaltenen Informationen kann das Vorhandensein der Schlüsselsequenz selbst Auskunft darüber geben, ob ein gekennzeichnetes Produkt echt ist, oder nicht. Die Information über die Authentizität des gekennzeichneten Produktes kann dabei über einen Prozeß, bei dem der Schlüsselprimer als Hybridisierungssonde einer Fluoreszenz-Nachweisreaktion dient, verifiziert oder falsifiziert werden.

Ein weiterer Gegenstand der vorliegenden Erfindung ist daher die Verwendung informationstragender Polymere, insbesondere der erfindungsgemäß erhältlichen informationstragenden Polymere, zur Verschlüsselung von Information.

Die oben beschriebenen Verfahren ermöglichen beispielsweise einen Test der Qualität von Oligonukleotiden: Ein typisches Problem bei der Herstellung von Oligonukleotiden, z. B. für molekularbiologische Zwecke, ist die Qualitätskontrolle derselben, die aufgrund des Herstellungsprozesses und der schwankenden Qualität der verwendeten Chemikalien problematisch ist.

Das in den erfindungsgemäßen Schritten I–V beschriebene Verfahren kann benutzt werden, die Qualität von Oligonukleotiden zu testen. Dazu wird eine binäre Grammatik wie die für den Zufallszahlengenerator (s. o.) oder eine unäre Grammatik wie die zur Herstellung von Molekulargewichtsstandards (s. u.) zur Erzeugung beliebig langer Logomere verwendet. Bei ansonsten identischen Bedingungen ist dann die Länge der aus einer Symbolpolymerisation (Schritt V des erfindungsgemäßen Verfahrens) erhaltenen Logomere direkt proportional zur Qualität der verwendeten Oligonukleotide: Je länger die elektrophoretisch sichtbar gemachten Logomere sind, desto besser ist die Qualität der verwendeten Oligonukleotide. Ein Beispiel für die aus einer Symbolpolymerisation erhaltene Leiter unterschiedlich langer Logomere findet sich unter Abb. 3.

Ein weiterer Gegenstand der vorliegenden Erfindung ist daher die Verwendung informationstragender Polymere, insbesondere der erfindungsgemäß erhältlichen informationstragenden Polymere, zur Qualitätskontrolle synthetisch hergestellter Oligonukleotide.

Die oben beschriebenen Verfahren ermöglichen beispielsweise die Herstellung von Logomeren als Marker und Signaturen:

Aufgrund der Fähigkeit, prinzipiell beliebige Informationen darzustellen, lassen sich die hier beschriebenen Logomere als Marker verwenden, um Erzeugnisse, Produkte, Stoffe und Geräte zu kennzeichnen.

Ein weiterer Gegenstand der vorliegenden Erfindung ist daher die Verwendung informationstragender Polymere, insbesondere der erfindungsgemäß erhältlichen informationstragenden Polymere, als Marker oder Signaturen.

Die Logomere können als "binäres Etikett" verwendet werden, das einem Erzeugnis hinzugefügt wird und Informationen über das derart gekennzeichnete Erzeugnis enthält. Die Logomere können dabei z. B. Informationen über

- a) Hersteller
- b) Produkt (z. B. Seriennummer)
- c) Verwendungszweck
- d) Stoffklasse
- e) Gefahrenklasse
- f) Qualität
- g) Reinheit
- h) Produktions- und Verfallsdatum

enthalten. Es lassen sich damit nahezu beliebige Produkte und Erzeugnisse mit nahezu beliebigen Informationen versehen. Da die Logomere ungiftig und biologisch abbaubar sind, können sie auch für hochsensible Produkte wie Nahrungsmittel, Medikamente und pharmazeutische Erzeugnisse eingesetzt werden.

Gekennzeichnete Erzeugnisse können z. B. sein:

- a) chemische Erzeugnisse, z. B. Lacke, Farben, Öle, Schmier- und Treibstoffe, Lösungsmittel, Tinte usw.
- b) flüssige Materialien: Lösungen, Suspensionen, Emulsionen
- c) gentechnische Erzeugnisse
- d) Nahrungsmittel, z. B. zur Kennzeichnung gentechnisch veränderter Bestandteile oder dem Monitoring von Nahrungsmitteln während des Produktionsprozesses
- e) Medikamente und pharmazeutische Erzeugnisse
- f) Papiererzeugnisse, Dokumente, Geld
- g) Geräte

Der Bedarf zur Kennzeichnung von Produkten besteht in den unterschiedlichsten Bereichen und aufgrund unterschiedlicher Anforderungen. Gründe für eine Kennzeichnung können z. B. sein: Qualitätssicherung beim Produktionsprozeß, Überwachung der Produktreinheit und Vermeidung von Kontaminationen, Schutz vor Fälschung und Produktpiraterie, Nachweis bestimmter Produktbestandteile, Produktkennzeichnung mit beliebigen Produktinformationen. Dieser Bedarf stellt sich auch in besonders sensiblen Bereichen, etwa bei der Kennzeichnung von Nahrungsmitteln, medizinischen und pharmazeutischen Erzeugnissen und gentechnisch hergestellten oder veränderten Erzeugnissen. Hier ist es wünschenswert, daß Informationen über das jeweilige Produkt direkt am Produkt verfügbar sind, insbesondere wenn

# DE 199 14 808 A 1

Verwechslungen oder Fälschungen vermieden werden sollen und auch die Identifikation unterschiedlicher Bestandteile eines Erzeugnisses gewünscht ist.

Es gibt zahlreiche Beispiele für den Bedarf von Produktkennzeichnungen. Z. B. können Logomere als Seriennummern eingesetzt werden, die in Autolacke gemischt werden, wodurch der Fahrzeughalter im Versicherungsfall (z. B. Unfall oder Diebstahl) identifiziert werden kann. Durch Produktkennzeichnungen könnte die Qualität und Reinheit chemischer Erzeugnisse überwacht und Kontaminationen vermieden werden. Ein permanentes Problem ist die Fälschung von Dokumenten, Unterschriften und Geld, die hochentwickelte Kennzeichnungsverfahren nötig macht.

Vor allem vor dem Hintergrund von Tierkrankheiten und -seuchen (BSE, Schweinepest, Salmonellenvergiftungen usw.) stellt sich das Problem der Qualitätssicherung von Nahrungsmitteln und der Kennzeichnung von Endprodukten, um kontaminierte Produkte vom Verbrauch fernzuhalten. Ein weiteres Problem sind Nahrungsmittel, die gentechnisch hergestellte Bestandteile enthalten. Diese Bestandteile können ohne ein Kennzeichnungsverfahren im Endprodukt überhaupt nicht oder nur sehr schwer nachgewiesen werden. Das Problem stellt sich insbesondere dadurch, daß viele Nahrungsmittel Bestandteile unterschiedlicher Herkunft enthalten und die Produktionswege teilweise undurchschaubar sind.

Das Problem der Kennzeichnung stellt sich z. B. auch im Bereich von Blutprodukten, wo z. B. durch Pooling Kontaminationen auftreten können. Auch wäre es hier z. T. wünschenswert, Informationen über Identität, Herkunft, Produktions- und Verfallsdatum direkt mit dem Produkt verfügbar zu haben. Ähnliches gilt auch für Medikamente und pharmazeutische Produkte.

Bisher verwendet man zur Kennzeichnung und Markierung von Farben, Lacken, Treibstoffen, usw. verschiedene Kohlenstoffverbindungen, Polyaniline, Flüssigkristalle oder andere chemische Verbindungen. Diese Stoffe und Verbindungen haben unterschiedliche Schwächen: sie sind z. T. toxisch, nur schwer biologisch abbaubar, die verfügbare Informationskapazität ist stark eingeschränkt, sie wechselwirken chemisch mit bestimmten Stoffen oder sind nur teuer und aufwendig herzustellen.

Für den Einsatz zur Kennzeichnung von sensiblen und hochsensiblen Erzeugnissen wie Nahrungsmitteln oder Medikamenten sind die genannten Stoffe und Verbindungen aufgrund der genannten Nachteile überhaupt nicht geeignet.

Die Anforderungen, die sich für ein Kennzeichnungsverfahren von beliebigen Erzeugnissen stellen umfassen zumindest:

- Die Kennzeichnung sollte direkt im oder am Erzeugnis verfügbar sein;
- die Kennzeichnung sollte gut nachweisbar sein;
- die Kennzeichnung muß gesundheitlich unbedenklich sein.

Gegenüber den bisher verwendeten Verbindungen haben die hier beschriebenen Logomere folgende Vorteile:

- Kodierung beliebiger Informationen,
- hohe Speicherkapazität,
- gesundheitlich unbedenklich,
- chemisch, biologisch, pharmakologisch und genetisch neutral,
- leicht biologisch abbaubar (Biomoleküle),
- keine Umweltbelastung,
- sehr gute Nachweisbarkeit,
- hohe Kompatibilität zu bestehenden Techniken der Informatik und Molekularbiologie (PCR),
- Informationen können verschlüsselt werden und eignen sich auch zu Authentifizierung und Verschlüsselung,
- Identifikation einzelner Bestandteile in Produktmischungen,
- billig in großen Mengen herzustellen (Klonierung, Vervielfältigung mit Bakterien in Fermentern).

Diese Eigenschaften machen Logomere für die genannten Aufgaben sehr viel besser geeignet als die bestehenden Techniken. Überdies erschließen sich Einsatzbereiche, die mit den bisherigen Techniken nicht möglich waren.

Aufgrund der Eigenschaften nukleinsäurebasierter Logomere können diese auch zur Kennzeichnung von besonders sensiblen Produkten wie Nahrungsmitteln und pharmazeutischen Erzeugnissen verwendet werden.

Die Gründe dafür liegen zum einen in der chemischen Beschaffenheit nukleinsäurebasierter Logomere und zum anderen in der Kodierung von Informationen in Logomeren:

- 1) Als Grundbausteine des Lebens" sind Nukleinsäuren Elementarbausteine aller bisher bekannten biologischen Lebensformen. Daher können sie aufgrund ihrer chemischen Beschaffenheit für keinen bekannten biologischen Organismus schädlich sein ("biochemische Kompatibilität"). Aufgrund der enthaltenen genetischen Informationen (des enthaltenen Programmes) können Nukleinsäuren jedoch sehr wohl potentiell schädlich sein, wenn sie in einem Organismus abgelesen werden: Falls sie etwa für giftige Proteine kodieren oder (wie im Falle viralen Erbgutes) einen jeweiligen Zielorganismus "umprogrammieren" können. Die hier verwendeten Logomere können jedoch aus den im Folgenden genannten Gründen keine für einen Organismus schädlichen Informationen enthalten.
- 2) Logomere enthalten keine genetischen, d. h. biologisch relevanten Informationen ("semantische Inkompabilität"): Bezuglich der enthaltenen Informationen entsprechen Logomere Buchstaben, Zahlen oder Folgen von Buchstaben und Zahlen in Binärkodierung, die für uns oder einen Computer, jedoch nicht für den genetischen Apparat eines Organismus lesbar sind. Für einen biologischen Organismus sind sie schlicht "Unsinnscode".
- 3) Logomere, insbesondere die Algomere, aus denen sie aufgebaut sind, sind zu kurz um auch nur zufällig biologisch relevante Informationen zu enthalten.
- 4) Um jede Eventualität (der unwahrscheinliche Fall, daß die Binärinformationen der Logomere auf irgendeine Weise von irgendeinem Organismus als genetisch relevante Information interpretiert wird) auszuschließen, werden die zur Kodierung von Symbolen verwendeten Sequenzen im Falle sensibler Anwendungsbereiche mit Stopcodons

# DE 199 14 808 A 1

versehen, so daß sie niemals von einem biologischen Organismus abgelesen werden können.

5) Als elementare Bestandteile aller biologischen Organismen sind Nukleinsäuren in unserer Umwelt allgegenwärtig. Sie werden in kürzester Zeit abgebaut.

Um als Marker eingesetzt zu werden, können Logomere nach den in den erfindungsgemäßen Schritten I-V beschriebenen Verfahren erzeugt und nach den erfindungsgemäßen oben beschriebenen Verfahren vervielfältigt werden. Je nach Einsatzbereich können zur Vervielfältigung der Logomere verschiedene Verfahren und verschiedene Aufarbeitungsprozesse vorgenommen werden. Zur Kennzeichnung petrochemischer Erzeugnisse z. B. können Logomere durch Klonierung in Bakterien vervielfältigt werden. Um unnötige Kontaminationen zu vermeiden, können die erhaltenen Bakterien lysiert werden. Eine spezielle Aufreinigung der Logomer-DNA ist im Allgemeinen nicht nötig.

Zur Herstellung von Logomeren für sensible und hochsensible Produkte ist dagegen ein aufwendigeres Vervielfältigungs- und Aufreinigungsverfahren nötig. Ziel ist dabei der Erhalt möglichst reiner Logomere. Um biologische Zwischenschritte zu vermeiden, kann die Vervielfältigung – wie erfindungsgemäß beschrieben – mit PCR vorgenommen werden. Sollte eine Vervielfältigung durch Klonierung vorgenommen werden, so ist die Aufreinigung der erhaltenen DNA aus Bakterien nötig. Sind die Logomere in bakteriellen Vektoren kloniert, so ist ein gezielter Abbau dieser Resistenzgene (z. B. durch Restriktionsenzyme) Bestandteil einer weiteren Aufarbeitung.

Logomere können zur Kennzeichnung eines Produktes unmittelbar mit diesem verbunden werden. Sie können z. B. in flüssige Stoffe gemischt werden oder auf feste Stoffe aufgebracht werden. Im Falle gentechnischer Erzeugnisse können die Logomere mithilfe von Rekombinations- und Klonierungstechniken kovalent mit dem zu kennzeichnenden Produkt verbunden werden.

Die Kennzeichnung von Stoffen und Produkten durch Logomere ermöglichen u. a. das Monitoring von Produktionsprozessen, die Qualitätskontrolle, die Identifikation einzelner Bestandteile und den quantitativen und qualitativen Nachweis von Kontaminationen.

Die oben beschriebenen Verfahren ermöglichen beispielsweise die Kennzeichnung gentechnisch hergestellter oder veränderter Erzeugnisse:

Die hier beschriebenen Logomere können zur Kennzeichnung gentechnisch hergestellter oder veränderter Erzeugnisse verwendet werden.

Bisher werden insbesondere zur Identifikation gentechnisch hergestellter oder veränderter Erzeugnisse in Nahrungsmitteln "klassische" molekularbiologische Methoden wie PCR, Restriktionsverdau, Southernblot-Hybridisierung und Sequenzierung verwendet. Mit diesen Methoden ist jedoch eine Charakterisierung gentechnischer Erzeugnisse und Bestandteile noch sehr aufwendig und erfordert für jedes zu charakterisierende Erzeugnis eigene Methoden und Verfahren.

Insgesamt gibt es bisher kein universelles Kennzeichnungsverfahren für gentechnisch hergestellte Erzeugnisse. Ein solches Kennzeichnungsverfahren muß mehrere Kriterien erfüllen:

- 35     – die Kennzeichnung muß gesundheitlich absolut unbedenklich sein,  
      – die Kennzeichnung sollte untrennbar mit dem gekennzeichneten Produkt verbunden sein,  
      – die Kennzeichnung sollte fälschungssicher sein,  
      – die Kennzeichnung sollte genügend Informationen speichern können,  
      – die Kennzeichnung sollte in geringen Mengen nachweisbar und lesbar sein.

40     Der Einsatz nukleinsäurebasierter Logomere zur Kennzeichnung gentechnischer Erzeugnisse erfüllt diese Kriterien und hat darüber hinaus folgende Vorteile:

- 45     – Logomere können nahezu beliebige Informationen speichern;  
      – einzelne Bestandteile können schnell und hochempfindlich nachgewiesen werden;  
      – zum Auslesen der Informationen kann eine bis ins Detail standardisierte Methode eingesetzt werden (Ausleseverfahren nach dem erfindungsgemäßen Verfahrens, dabei können standardisierte Primer verwendet werden);  
      – die Sequenz der gesuchten Gene kann vollkommen unbekannt sein;  
      – die Sequenz der nachgewiesenen Gene kann geheim sein, ohne daß die Identifizierung oder Authentifizierung eingeschränkt ist;  
      – zusätzliche Sicherheitsmechanismen können zusätzlich durch Verschlüsselung implementiert werden.

55     Im Einzelnen geht man zur Kennzeichnung von gentechnisch erzeugten oder veränderten Erzeugnissen beispielsweise folgendermaßen vor:

55     Zur Herstellung von Markern wird eine beliebige geeignete Grammatik, z. B. wie die für den Zufallszahlengenerator, das 1-Byte Alphabet oder für Zeichenketten bereits beschriebene, verwendet. Aus den mit der Grammatik erzeugten Zeichen werden die zur Darstellung von Markern benötigten Zeichen entnommen und dem zu kennzeichnenden Erzeugnis hinzugefügt. Dazu können die als Logomere hergestellten Zeichen auf folgende Weise in das zu kennzeichnende Erzeugnis eingebbracht werden:

- 60     a) durch Beimischung; hierbei werden Logomere, wie in den erfindungsgemäßen Verfahren beschrieben, hergestellt, isoliert und vervielfältigt.  
      b) durch Klonierung, wie im erfindungsgemäßen Verfahren beschrieben; Abb. 6 zeigt z. B. die Kennzeichnung bakterieller Klone durch Logomere, die durch Verwendung der bereits beschriebenen Grammatik für den Zufallszahlengenerator hergestellt wurden.  
      c) durch den Einsatz rekombinativer Techniken; hierbei werden Logomere erfindungsgemäß hergestellt, isoliert und vervielfältigt. Die solcherart gewonnenen Logomere werden jetzt durch den Einsatz rekombinativer Techniken in das zu kennzeichnende Erzeugnis eingebracht. Dazu kann z. B. die Methoden des Gene-Targeting durch homo-

logische Rekombination [Capecci M. R., Altering the genome by homologous recombination. *Science*, 244(4910), 1288–1292, (1989)] oder z. B. das Cre-loxP System [Kilby, N. J., Snaith, M. R. & Murray, J. A., Site-specific recombinases: tools for genome engineering, *Trends Genet.*, 9, 413–421, (1993)] verwendet werden. Z. B. können erfundungsgemäß erhaltene Logomere mit Hilfe dieser Techniken vor oder hinter einer bestimmten Nukleinsäuresequenz (Gen) gesetzt werden, die sie bezeichnen sollen. In diesem Falle wird ein Logomer gewissermaßen als "Etikett" an ein gentechnisches Produkt "angeheftet" und kann Auskunft über Hersteller, Produktgruppe, Gefahrenklasse, Verfallsdatum o. ä. geben (siehe Abb. 15). 5

Das Verfahren eignet sich z. B. für alle Organismen (Mikroorganismen, Pflanzen, Tiere). Z. B. lassen sich Klone von Mikroorganismen oder Pflanzen kennzeichnen. Das Verfahren eignet sich z. B. auch für die Kennzeichnung von Rindern zur Bekämpfung der Rinderseuche BSE. 10

Die oben beschriebenen Verfahren ermöglichen beispielsweise die Kennzeichnung von Nahrungsmitteln: Aufgrund von Krankheiten, Seuchen oder Kontaminationen können Nahrungsmittel gesundheitsschädlich oder gefährlich sein. Beispiele sind Krankheiten wie BSE, Schweinepest, Salmonellenvergiftungen. Es kann wünschenswert sein, Produkt und Hersteller identifizieren zu können, um notfalls bestimmte Produkte und Produktreihen umgehend aus dem Handel zu nehmen, untersuchen zu können und geeignete Gegenmaßnahmen zu ergreifen. 15

Für hochsensible Erzeugnisse wie Nahrungsmittel gelten dabei die Kriterien für die Kennzeichnung gentechnischer Erzeugnisse insbesondere (siehe: Die oben beschriebenen Verfahren ermöglichen beispielsweise die Kennzeichnung gentechnisch hergestellter oder veränderter Erzeugnisse). 20

Insgesamt gibt es auch für gentechnisch hergestellte Nahrungsmittel bisher kein universelles Kennzeichnungsverfahren. Bisher ist die Identifikation gentechnisch hergestellter oder veränderter Bestandteile in Nahrungsmitteln daher noch schwierig und aufwendig. Es werden dazu "klassische" molekularbiologische Methoden wie PCR, Restriktionsverdau, Southernblot-Hybridisierung und Sequenzierung verwendet, um diese Erzeugnisse zu charakterisieren. 25

Aufgrund der Rinderseuche BSE wurde speziell für die Kennzeichnung von Rindern und deren Produkten (Rindfleisch und Milch) ein immunologisches Kennzeichnungsverfahren vorgeschlagen, das auf der Immunisierung der zu markierenden Tiere mit spezifischen Proteinen basiert ("Ohrmarke im Blut soll BSE Rinder aufspüren", Süddeutsche Zeitung Nr. 283, 08.12.1998, Seite v2/9). Durch die dadurch verursachte Produktion spezifischer Antikörper im jeweils gekennzeichneten Zieltier, die mutmaßlich in allen Geweben nachzuweisen ist, kann die Kennzeichnung über einen Immun-Assay (ELISA) wieder gelesen werden. Ein solches Verfahren ist jedoch prinzipiell auf höhere Wirbeltiere beschränkt. Außerdem ist die für das Verfahren verfügbare Informationsmenge sehr begrenzt, die informationstragenden Proteine sind nicht hitzebeständig und das Verfahren erfordert vergleichsweise hohe Mengen Substanz für Immunisierung und Nachweisreaktion. 30

Auch im Fall von Nahrungsmitteln können Logomere auf Nukleinsäurebasis, wie bereits unter "Die oben beschriebenen Verfahren ermöglichen beispielsweise die Kennzeichnung gentechnisch hergestellter oder veränderter Erzeugnisse" beschrieben, zur Kennzeichnung eingesetzt werden. Logomere können dabei Informationen über Hersteller, Verfallsdatum, Seriennummer u. a. enthalten. Das Verfahren zur Kennzeichnung von Nahrungsmitteln ist dabei dasselbe, wie bereits unter "Die oben beschriebenen Verfahren ermöglichen beispielsweise die Kennzeichnung gentechnisch hergestellter oder veränderter Erzeugnisse" beschrieben. Mithilfe der dort aufgeführten gentechnischen Rekombinations- und Klonierungstechniken können einzelne Organismen gekennzeichnet werden, sodaß ein einzelner Organismus sowie alle seine Nachfahren identifizierbar sind. 40

Das Kennzeichnungsverfahren durch Logomere ermöglicht dadurch:

- ein Monitoring des Zielproduktes über den gesamten Herstellungs- und Verarbeitungsprozesses bis zum Endverbraucher,
- eine einheitliche, schnelle und einfache Identifikation einzelner Bestandteile. 45

Als Beimischung eignen sich Logomere auf Nukleinsäurebasis zur Kennzeichnung von Getränken.

Die oben beschriebenen Verfahren ermöglichen beispielsweise die Kennzeichnung medizinischer und pharmazeutischer Produkte: 50

Aufgrund der Nichttoxizität eignen sich Logomere auf Nukleinsäurebasis zur Kennzeichnung sensibler und hochsensibler Erzeugnisse. Neben Nahrungsmitteln sind dies auch medizinische und pharmazeutische Produkte. Um Verwechslungen, Kontaminationen und Fälschungen ausschließen zu können, können Logomere z. B. medizinischen und pharmazeutischen Produkten beigemischt werden. 55

Der Vorteil der Kennzeichnung durch Logomere ist dabei, daß die Kennzeichnung unmittelbar mit dem gekennzeichneten Erzeugnis verbunden ist und daß sich Erzeugnisse z. B. individuell kennzeichnen lassen. Auf diese Weise ist ein Monitoring des Herstellungsprozesses möglich, Erzeugnisse lassen sich auch nach mehrmaligem Umladen eindeutig identifizieren und Kontaminationen sowie das Poolen von Proben können nachgewiesen werden. Mithilfe von Logomen läßt sich z. B. im Falle von Blutkonserven Herstellungs- und Verfallsdatum, Blutgruppe und Hersteller kennzeichnen. 60

Die oben beschriebenen Verfahren ermöglichen beispielsweise Authentifizierung und Fälschungssicherung: Logomere können zur Authentifizierung von Erzeugnissen, Gegenständen und Geräten verwendet werden. Dazu werden Logomere vorzugsweise erfundungsgemäß erzeugt und vervielfältigt. 65

Werden solcherart erhaltene Logomere auf Erzeugnisse, Gegenstände und Geräte an-/aufgebracht oder zugemischt, so können die in Logomeren enthaltenen erfundungsgemäß jederzeit wieder ausgelesen werden. Sollten zur Erzeugung der Logomere Verschlüsselungstechniken verwendet werden, wie bereits beschrieben, so erfordert das Auslesen autorisierten Zugriff. 70

Beispielsweise können erfundungsgemäß erhaltene Logomere in wässriger Lösung (vorzugsweise zu  $10^3$  bis  $10^{18}$  Molekülen/ $\mu$ l) unmittelbar zur Kennzeichnung auf Dokumente aufgebracht werden. Als solches können sie als Echtheitszertifikat dienen. 75

# DE 199 14 808 A 1

tifikat des solcherart markierten Dokumentes dienen. Zum Auslesen der als Kennzeichner dienenden Logomere wird einfach ein kleines Schnipsel des Papiers ( $1 \text{ mm}^2$  ist ausreichend) als Template einer erfundungsgemäßen PCR-Auslesereaktion verwendet. Ein Beispiel findet sich in Abb. 16.

In dem zugrundeliegenden Ausleseexperiment wurde zum Test des Prinzips ein Logomer in wässriger Lösung in ca. 5  $10^9$  Molekülen/ $\mu\text{l}$  auf 3M Postlt Papier ca. 1 Stunde getrocknet und  $1 \text{ mm}^2$  des Post-lts als Template einer Auslese-PCR verwendet. Es können Dokumente beliebiger Papierqualität verwendet werden, z. B.  $80 \text{ g/m}^2$  Standardkopierpapier.

Dasselbe Verfahren eignet sich auch zur Kennzeichnung von Geld oder Geldscheinen, wobei hier Vorteile darin bestehen,

- 10     – auch bereits gedruckte Scheine markieren zu können,
- Seriennummern vergeben zu können,
- Verschlüsselung einzusetzen.

Um Dokumente möglichst in Echtzeit auf Authentizität prüfen zu können, kann einer der zum Auslesen in der PCR benutzten Primer auch als Hybridisierungsonde für eine nachfolgende Färbe-Nachweisreaktion (z. B. mit einem interkalierenden Farbstoff, z. B. Ethidiumbromid) verwendet werden.

Ein weiteres Beispiel ist die Authentifizierung flüssiger Lösungen, Suspensionen und Emulsionen mit Logomeren. So kann etwa Tinte mit Logomeren individuell markiert werden, so daß Unterschriften damit auf Echtheit überprüft werden können.

20     Ein weiteres Beispiel ist die Kennzeichnung von Automobilen durch das Versetzen des Autolackes mit Logomeren. Dabei werden Logomere dem Autolack (oder anderen Bestandteilen des KFZ) als Beimischung zugesetzt. Die zugesetzten Logomere können dabei Informationen über Seriennummer, Hersteller, Fahrzeugtyp o. a. Angaben enthalten. Aufgrund dieser Informationen kann im Falle von Unfall, Diebstahl oder illegaler Entsorgung ein Fahrzeug identifiziert werden. Ein Vorteil der hier beschriebenen Methode liegt darin, daß schon Lackspuren zur Identifikation des Fahrzeugs ausreichen, was insbesondere bei Unfällen von Interesse sein kann. Ein weiterer Vorteil ist, daß die Identität eines Fahrzeuges nur sehr aufwendig zu fälschen ist. Dazu müßten erstens alle gekennzeichneten Bestandteile entfernt, zweitens eine gefälschte Identität eingerichtet werden. Dieses Unterfangen ist alleine durch den mechanischen Aufwand sehr schwierig, außerdem läßt sich durch den Einsatz kryptografischer Methoden das Nachmachen und Nachahmen von Kennungen beliebig erschweren. Zum Auslesen der als Logomere enthaltenen Informationen, werden die Logomere aus dem Lack isoliert und können dann wie in der Beschreibung zum Auslesen der in Logomeren enthaltenen Information, vorzugsweise durch PCR, ausgelesen werden. Dazu wird eine Probe des Lackes entnommen, mit einem geeigneten Lösungsmittel (Terpentin o. ä.) gelöst und mit gleichem Volumen Wasser versetzt. Durch nachfolgende Zentrifugation werden hydrophile und hydrophobe Bestandteile getrennt. Daraufhin nimmt man die wässrige Phase auf (worin sich die hydrophilen nukleinsäurebasierten Logomere befinden) und fällt die enthaltenen Logomere nach den üblichen Methoden (z. B. wie in [J. Sambrook, E. F. Fritsch, T. Maniatis, Molecular Cloning, A Laboratory Manual, (1989)] beschrieben). Nach der Fällung können die erhaltenen Logomere nach den erfundungsgemäßen Verfahren, z. B. durch Auslese-PCR, ausgelesen werden.

Daher sind weitere Gegenstände der vorliegenden Erfindung die Verwendungen informationstragender Polymere, insbesondere der erfundungsgemäß erhältlichen informationstragenden Polymere, zum Zwecke der Qualitätssicherung, der Fälschungssicherung, der Kennzeichnung von Lebensmitteln, der Kennzeichnung gentechnischer, chemischer, medizinischer und pharmazeutischer Erzeugnisse, der Kennzeichnung von Organismen, der Kennzeichnung von Dokumenten, der Kennzeichnung von Geld, der Kennzeichnung von Gegenständen und Maschinen, der Kennzeichnung von Lösungen, Suspensionen und Emulsionen sowie der Authentifizierung von Personen und Gegenständen.

Die oben beschriebenen Verfahren ermöglichen beispielsweise die Herstellung von Molekulargewichtsstandards:

45     Zu den typischen Verfahren der molekularbiologischen Analytik gehört die elektrophoretische Längenauf trennung von DNA- und RNA-Fragmenten. Um die Länge solcher Fragmenten bei der Elektrophorese zu bestimmen, verwendet man üblicherweise einen Molekulargewichtsstandard ("Leiter"), der Fragmente bekannter Länge enthält und zum Vergleich mit den Fragmenten der zu analysierenden Proben dient.

Bisher gibt es Molekulargewichtsstandards, die entweder unregelmäßige Leitern enthalten (z. B. Boehringer V, Boehringer Mannheim, Katalog Nr.: 821 705) oder regelmäßige Leitern mit regelmäßigen Fragment-Längenabständen (z. B. 50 bp Leiter Gibco BRL, Life Technologies, Katalog Nr. 10416-014). Keine der bisherigen Leitern hat dabei kürzere Längenabstände als 10 bp (z. B. 10 bp Leiter Gibco BRL, Life Technologies, Katalog Nr. 10821-015).

Mit den hier beschriebenen Verfahren können gezielt DNA Fragmente verschiedener Länge und definierter Längenabstände hergestellt werden. Diese Fragmente können als Molekulargewichtsstandards genutzt werden, wobei das Verfahren erlaubt, prinzipiell beliebige Längen und beliebige Längenabstände zu realisieren. Insbesondere sind Längenabstände bis hinab zu 1 bp Abstand, also Molekulargewichtsstandards höchster Auflösung, möglich.

Das Verfahren beruht darauf, Logomere als Template einer Auslese-PCR zu verwenden, deren Resultat ein Gemisch von DNA Fragmenten definierter Länge und definierter Längenabstände ist. Es basiert auf den bereits beschriebenen erfundungsgemäßen Verfahren zur Erzeugung und dem Auslesen von Logomeren.

60     DNA Fragmente verschiedener Länge in sehr kurzen, regelmäßigen Abständen können erhalten werden, wenn ein unäres Polymer als Template einer Auslese-PCR mit verschachtelten Primern fungiert. Der 5'-Primer pri mt in der Start-Sequenz des Polymers. Als gegenläufige 3' Primer fungiert eine Gruppe von Primern, die verschachtelt in dem Elongator ansetzt (in Abb. 17 dargestellt als 3 Ebenen gegeneinander versetzter Primer). Da das Polymer aus Wiederholungen nur eines Elongator-Moleküls besteht, erhält man für  $n$  Wiederholungen und  $m$  verschachtelte 3' Primer  $n \cdot m$  Banden. Das Verfahren funktioniert auch spiegelbildlich mit einem im Ende-Terminator primenden Primer und dementsprechend gegenläufigen 5' Primern. Die erhaltenen Banden können als Molekulargewichtsstandards z. B. in Elektrophoresen eingesetzt werden.

Für hochauflösende Molekulargewichtsstandards ist die Verwendung unärer Logomere und verschachtelter Primer am

# DE 199 14 808 A 1

zweckmäßigsten. Jedoch ist das Verfahren nicht darauf beschränkt, da im Prinzip alle im erfindungsgemäßen Schritt I beschriebenen Grammatiken verwendet werden können.

Unäre Logomere beliebiger Länge können mit einer Grammatik  $G = (\Sigma, V, R, S)$  mit Terminalalphabet  $\Sigma = \{0, s, e\}$ , Variablenmenge  $V = \{A\}$ , Startsymbol  $S$  und Regelmenge

```
R :=  
{  
S := sA  
A → aA  
A → e  
}
```

5

hergestellt werden. Sollte es nötig sein, unäre Logomere definierter Länge herzustellen, muß dementsprechend die Variablenmenge und damit auch die Zahl der Elongatoren größer gewählt werden (wie z. B. in der Grammatik zur Implementierung eines 1-Byte Alphabets bereits beschrieben). Je nach gewünschter Länge der zu erzeugenden DNA Fragmente kann außerdem die Länge der Algomere variiert werden.

Die so erzeugten Logomere können nach den erfindungsgemäßen Verfahren vereinzelt und vervielfältigt werden.

Das Erzeugen von DNA-Fragmenten definierter Länge und definierter Längenabstände erfolgt vorzugsweise mit PCR analog dem erfindungsgemäßen Verfahren zum Auslesen von Logomeren. Für hochauflösende Molekulargewichtsstandards wird pro Elongator mindestens ein Primer, meistens sogar eine Anzahl mehrerer Primer verwendet, die gegeneinander verschoben sind. Die Anzahl der Primer pro Elongator ist abhängig von den gewünschten Längenabständen. Ist etwa gewünscht, Algomere der Länge 30 bp zu verwenden und sollen die Längen der erzeugten DNA Fragmente jeweils 10 bp Unterschied haben, so sind pro Elongator drei Primer, die jeweils genau 10 bp gegeneinander verschoben sind, eingesetzt werden.

10

Für die Synthese von Sequenzen zur Herstellung von Algomenen und Logomenen müssen die Anforderungen erfüllt werden, die bereits unter Schritt II des erfindungsgemäßen Verfahrens beschrieben wurden. Insbesondere gewährleistet das verwendete NFR-Verfahren dadurch, daß die jeweils als Template dienenden Teilsequenzen einen gleichen GC-Anteil besitzen können, den Einsatz verschachtelter Primer und die Ausgewogenheit des AT/GC Verhältnisses der erzeugten DNA Fragmente. Letzteres ist insbesondere angesichts dessen entscheidend, daß das Laufverhalten von DNA Fragmenten in der Elektrophorese nicht nur von der Länge, sondern auch von der Zusammensetzung der Fragmente abhängt. Dies gilt insbesondere für hochauflösende Molekulargewichtsstandards in denen es nur sehr geringe Längenabstände gibt.

15

Ein weiterer Gegenstand der vorliegenden Erfindung ist daher die Verwendung informationstragender Polymere, insbesondere der erfindungsgemäß erhältlichen informationstragenden Polymere, zur Herstellung von Molekulargewichtsstandards.

20

Die oben beschriebenen Verfahren ermöglichen beispielsweise die Herstellung polymerer Datenspeicher: Die beschriebenen Logomere können als RAM (Lesen und Schreiben) und ROM (nur Lesen) Datenspeicher hoher Kapazität und niedrigen Energieverbrauchs genutzt werden. Um eine möglichst hohe Informationsdichte zu gewährleisten und die Handhabung des Speichers möglichst einfach zu machen, werden die Logomere an einen festen Träger gebunden. Dies gewährleistet, daß die Logomere einzeln adressiert, d. h. einzeln geschrieben, gelesen und gelöscht werden können (siehe Abb. 9).

25

Die benötigten Schreib- und Leseoperationen werden auf der Basis der im erfindungsgemäßen Schritt V und der unter Auslesen der in Logomenen enthaltenen Information beschriebenen enzymatischen Reaktionen durchgeführt, die ihrerseits durch Temperaturänderungen gesteuert werden. Die Temperaturänderungen können entweder per Laser oder durch Erhitzen und Abkühlen des festen Trägers z. B. in einem Thermozyklus durchgeführt werden.

30

Zum Schreiben wird eine Modifikation des oben beschriebenen Verfahrens der Symbolpolymerisation verwendet. Die Polymerisation findet an einem festen Träger statt, die Algomere werden durch wiederholte Restriktions-Ligationszyklen miteinander verkettet.

45

Um einzelne, adressierbare Speicherzellen herzustellen, werden in bestimmten Abständen einzelsträngige "Ankermoleküle" irreversibel (durch UV-Bestrahlung oder im Ofen) an den Träger gebunden. Der solcherart mit Ankermolekülen versehene Träger kann nun reversibel mit Logomenen beschrieben werden, die in einem Löschvorgang wieder davon entfernt werden können.

50

Bei einem Schreibvorgang werden zuerst spezifische Algomer ("Starter-Algomer") durch Hybridisierung an die Ankermoleküle gebunden. Diese Starter-Algomere sind ihrerseits Ausgangspunkt einer Symbolpolymerisation, bei der weitere Algomere miteinander verketten werden. Die Verkettung der Algomere erfolgt dabei, abweichend von dem im Schritt V des erfindungsgemäßen Verfahrens beschriebenen Vorgehen, durch wiederholte Zyklen von Restriktion und Ligation. Bei jedem Zyklus wird das jeweils letzte Algomer einer polymerisierenden Kette mit einem Restriktionsenzym geschnitten, so daß eine einzelsträngige Überhangsequenz (Elongationspunkt) entsteht, an die seinerseits das nächstfolgende Algomer durch Ligation gebunden werden kann.

55

Für den solcherart beschriebenen Schreibvorgang müssen die Algomere ein spezifisches Design aufweisen:

- a) Jedes zu schreibende Algomer besitzt eine einzelsträngige Überhangsequenz, mit der es an den Elongationspunkt eines Logomers binden kann.
- b) Ein Algomer, das nur eine einzelsträngige Überhangsequenz besitzt mit der es an einen Elongationspunkt binden kann, selbst aber keine Restriktionsschnittstelle besitzt, durch die es von Restriktionsenzymen erkannt werden kann und damit nicht als Elongationspunkt dienen kann, heißt Ende-Algomer.
- c) Jedes zu schreibende Algomer, das nicht das Ende-Algomer ist, besitzt ein doppelsträngiges Ende, worin sich die Erkennungssequenz für das jeweils gewählte Restriktionsenzym befindet. Bei einer Restriktion spaltet das Enzym das Algomer, so daß eine einzelsträngige Überhangsequenz entsteht, die ihrerseits selbst wieder als Elongationspunkt dienen kann.
- d) Jedes fertige Logomer besitzt genau ein Starter- und ein Ende-Algomer. Starter- und Ende-Algomere sind, ent-

60

65

# DE 199 14 808 A 1

sprechend der oben gegebenen Definition, Terminatoren.

e) Die Überhangsequenzen der Algomere sind kompatibel zu einem jeweils gewählten Restriktionsenzym, so daß ein zu verkettendes Algomer an den Elongationspunkt eines Logomers binden kann und der nachfolgend durch Restriktion entstehende Elongationspunkt identisch zu seinem Vorläufer ist.

5 f) Die im Algomer auf die Überhangsequenz folgende Sequenz ist solcherart gewählt, daß ein verkettetes Algomer durch einen nachfolgenden Restriktionsvorgang nicht wieder abgespalten werden kann.

Als fester Träger eignen sich Membranen wie Gene Screen Plus (DuPont, Biotechnology Systems, Katalog Nr.: NEF-986 oder NEF-987), Hybond-N (Amersham Life Sciences, Katalog Nr.: RPN 82N oder RPN 137 N), Silizium-, Silicat-10 Oberflächen oder Glas. Die Ankermoleküle werden jeweils kovalent daran gebunden (UV-Bestrahlung oder Hitze).

Damit die Moleküle einzeln adressierbar sind, müssen sie in regelmäßigen Abständen auf dem Träger angebracht werden.

Als Restriktionsenzyme eignen sich handelsübliche Enzyme, vorzugsweise solche, die nicht bei 65°C deaktiviert werden (z. B. HindIII).

15 Die Schreiboperationen basieren darauf, daß die zum Schreiben verwendeten Enzyme unterschiedliche Temperaturoptima haben. So liegt das Temperaturopimum der zum Verketten von Symbolen verwendeten Ligase bei 16°C, wohingegen die benötigten Restriktionsenzyme nur bei 37°C funktionieren. Dadurch ist das getaktete Schreiben von Symbolen in Zyklen von Restriktion und Ligation möglich.

Da alle Lese- und Schreiboperationen mithilfe von Enzymen durchgeführt werden, ist es nötig, den Träger temperieren 20 zu können. Dabei können, bei der Verwendung eines Thermozyklers und thermomanipulierbaren Materials, alle Speicherzellen simultan derselben Operation unterworfen werden. Alternativ können Speicherzellen einzeln und unabhängig voneinander zugegriffen werden, wenn ein entsprechend kleiner Schreib-/Lesekopf eingesetzt wird. Dieser Schreib-/Lesekopf muß einerseits als Pipettierzvorrichtung für die benötigten Moleküle (Enzyme, Algomere) dienen und andererseits die für eine bestimmte Manipulation benötigte Temperatur erzeugen (z. B. mit Hilfe eines Lasers).

25 Beim Löschvorgang wird ein Logomer durch Denaturierung vom jeweiligen Ankermolekül gelöst. Das Ankermolekül steht danach wieder für einen Schreibvorgang zur Verfügung. Für einen Lesevorgang wird ein Logomer durch Denaturierung von einem Ankermolekül gelöst. Es kann dann nach den erfundungsgemäßen Verfahren ausgelesen werden.

Die beschriebenen Operationen Schreiben, Löschen, Lesen werden durch Enzyme bei verschiedenen Temperaturen 30 durchgeführt. Dabei können, bei der Verwendung eines thermomanipulierbaren Materials, alle Speicherzellen simultan derselben Operation unterworfen werden. Alternativ können Speicherzellen einzeln und unabhängig voneinander zugegriffen werden, wenn ein entsprechend kleiner Schreib-/Lesekopf eingesetzt wird. Dieser Schreib-/Lesekopf muß einerseits als Pipettierzvorrichtung dienen und andererseits die für eine bestimmte Manipulation benötigte Temperatur erzeugen.

Der Vorteil des hier vorgestellten Polymerspeichers besteht gegenüber den bisher verwendeten Materialien in einer 35 höheren Speicherdichte (DNA Moleküle haben eine Größe, die bei  $10^{-10}$  m liegt), und einem sehr viel geringeren Energieverbrauch (zur Berechnung der in enzymatischen Reaktionen benötigten Energie siehe [L. M. Adleman, Molecular Computation of Solutions to Combinatorial Problems, Science, 266, 1021–1024, (1994)]). Außerdem ergibt sich eine erheblich verbesserte Umweltverträglichkeit, weil die hier beschriebenen Polymere Biomoleküle ohne jegliche Toxizität sind.

40 Der hier beschriebene Ansatz hat gegenüber den bisher für das DNA Computing verwendeten wässrigen Lösungen den Vorzug höherer Speicherdichte und der Adressierbarkeit einzelner Polymere.

Ein weiterer Gegenstand der vorliegenden Erfindung ist daher ein polymerer Datenspeicher, umfassend erfundungsgemäße informationstragende Polymere, erfundungsgemäße Verfahren zur Verknüpfung von Molekülen, erfundungsgemäße Ausleseverfahren oder erfundungsgemäße Isolierungs- und Vervielfältigungsverfahren.

45 Die oben beschriebenen Verfahren ermöglichen beispielsweise die Herstellung eines DNA Computers: Mithilfe der bereits beschriebenen Informationsdarstellung durch Algomere und Logomere kann ein DNA Computer konstruiert werden. Der DNA Computer umfaßt (s. Abb. 10):

- 50 a) Oligonukleotidsynthesizer
- b) Thermozykler
- c) Pipettierzvorrichtung
- d) Vorrichtung zur Isolation von Polymeren
- e) Vorrichtung zur Auftrennung von Nukleinsäuren, z. B. Elektrophoresesystem
- f) Scanner
- 55 g) Steuerungsrechner (Workstation, z. B. PC)

Der DNA Computer ist konstruiert als ein Hybridsystem, bei dem einfache, rechenintensive Algorithmen oder die Speicherung von Daten mithilfe von DNA und den dazu benötigten Geräten (Geräte a) bis e)) vorgenommen werden und ein herkömmlicher Computer (vorzugsweise Workstation, z. B. PC) als Host und Steuerungsrechner dient.

60 Der DNA Computer beinhaltet mindestens einen Thermozykler, der als "Informationsreaktor" dient und über den Steuerungsrechner gesteuert werden kann (z. B. MJ Research PTC-100, Eppendorf Mastercycler). In den Tubes oder Mikrotiterplatten des Thermozyklers befinden sich Molekülgemische aus Nukleinsäuren, Enzymen und weiteren chemischen Substanzen (z. B. Puffer, Nukleotide), die für die Algomer-Assemblierung (Schritt IV des erfundungsgemäßen Verfahrens) und die Symbolpolymerisation (Schritt V des erfundungsgemäßen Verfahrens) benötigt werden. Programmiert wird der DNA Computer durch die Implementierung von Algorithmen als Grammatiken (Schritt I des erfundungsgemäßen Verfahrens).

65 Die für die Grammatiken jeweils benötigten Monomersequenzen werden durch das NFR-Verfahren und die Benutzung eines Oligonukleotidsynthesizers (z. B. ABI 392, ABI 398, ABI 3948 von Perkin-Elmer Applied Biosystems) hergestellt (Schritte II und III des erfundungsgemäßen Verfahrens). Diese Monomersequenzen gehen als Input

# DE 199 14 808 A 1

in den Thermozykler ein, in dem die Algomer-Assemblierung (Schritt IV des erfindungsgemäßen Verfahrens) stattfindet.

Die so gewonnenen Algomere werden in eine oder mehrere Reaktionskammern des Thermozyklers überführt, wo sie zu Logomeren verknüpft werden (Symbolpolymerisation, Schritt V des erfindungsgemäßen Verfahrens). Die so gewonnenen Logomere werden in einer eigenen Vorrichtung isoliert und vervielfältigt. Diese Vorrichtung kann nach den Prinzipien arbeiten, die in dem erfindungsgemäßen Verfahren beschrieben worden sind. Nach Isolierung und Vervielfältigung können die Logomere gelesen werden. Dazu werden sie wiederum in einem Thermozykler erfindungsgemäß ausgelesen. Die aus dem Ausleseverfahren resultierenden Nukleinsäurefragmente können, wie bereits beschrieben, durch Gelektrophorese gelesen werden. Hier dient das Gelektrophoresegerät als Ausgabegerät des DNA Computers, die solcherart erhaltenen Bandenmuster werden von einem Scanner in den Steuerungsrechner eingelesen.

Der Steuerungsrechner nimmt jetzt seinerseits aufgrund der solcherart erhaltenen Ergebnisse weitere Berechnungen und Steuerungen vor und setzt ggfs. weitere molekulare Reaktionen in Gang. Z. B. kann der Steuerungscomputer aufgrund erhaltener Ergebnisse die Neuimplementierung von Grammatiken (Herstellung von Oligomeren mit dem NFR-Verfahren) veranlassen. Auf diese Weise können die durch molekulare Verfahren als Logomere erhaltenen Informationen nicht nur als passive Daten sondern auch als Instruktionen (und Adressen) dienen. Dadurch ist ein sich selbst steuernder Prozeß verschiedener Algorithmen möglich, der für eine universelle Datenverarbeitung nötig ist.

Der DNA Computer verarbeitet dabei Informationen als Oligo- und Polymere. Als Speicherzellen der Daten dienen Klonierungsvektoren (Plasmide), die sich in flüssiger Lösung oder auf einem festen, adressierbaren Träger befinden.

Der DNA Computer ist wie oben beschrieben mittels Grammatiken programmierbar und kann von einem herkömmlichen Computer, der als Steuerungsrechner und Host-System fungiert, gesteuert werden. Da der DNA Computer bestimmte Operationen in molaren Größenordnungen (z. B. die Symbolpolymerisation, wie in Schritt V des erfindungsgemäßen Verfahrens beschrieben, in  $10^{12}$ – $10^{20}$  Elementaroperationen pro Sekunde) vornehmen kann, ist eine mögliche Anwendung die Berechnung einfacher, rechenintensiver Algorithmen. Jedoch kann der DNA Computer auch für andere Anwendungen, z. B. die Erzeugung bestimmter, regelmäßiger DNA Strukturen, die z. B. innerhalb der Nanotechnologie interessant sind, verwendet werden.

Ein weiterer Gegenstand der vorliegenden Erfindung ist daher ein DNA-Computer, umfassend erfindungsgemäße informationstragende Polymere. Ein weiterer Gegenstand der vorliegenden Erfindung ist ein DNA-Computer, in dem ein erfindungsgemäßes Ausleseverfahren und/oder erfindungsgemäßes Isolierungs- und Vervielfältigungsverfahren eingesetzt werden.

Die oben beschriebenen Verfahren ermöglichen beispielsweise die Herstellung kleinster molekularer Strukturen mit Logomeren:

Mit den hier beschriebenen Logomeren lassen sich kleinste molekulare Strukturen (Nanostrukturen) herstellen. Bausteine solcher Strukturen sind Algomere, die zu regelmäßigen Mustern von Logomeren zusammengefügt werden können. Z. B. können verschiedene Algomere mit unterschiedlichen Fremdmolekülen (Liganden) dotiert werden, um regelmäßige, höhermolekulare Muster der jeweiligen Liganden zu erhalten. Die Logomere fungieren dabei als "Skelett" molekularer Baugruppen von Liganden.

Z. B. könnte ein binäres Logomer alternierende Muster leitfähiger und nicht-leitfähiger Liganden enthalten. Durch die Anordnung vieler Logomere auf einem festen Träger können so Leiterbahnen im Nanometerbereich entstehen.

Als Liganden können aber auch Biomoleküle, Antikörper, optisch aktive Moleküle o.ä. verwendet werden.

Logomere können als "intelligenter" Klebstoff eingesetzt werden (siehe Abb. 18). Dabei macht man sich die Hybridisierung komplementärer Nukleotide zunutze. Bringt man auf die zu verklebenden Flächen Logomere auf, die etwa kovalent an die zu verklebenden Flächen gebunden werden, so können ansonsten völlig identische Flächen nur dann verkleben, wenn sie komplementäre Sequenzen aufweisen, da nur jeweils Flächen komplementärer Nukleinsäuren Wasserstoffbrückenbindungen ausbilden. Die "Intelligenz" des Klebstoffes besteht also in einer durch die jeweils verwendeten Sequenzen hochselektiven Adhäsionswirkung. Dabei kann auch die Adhäsionsstärke für zu verklebende Flächen genau eingestellt werden: Zum einen kann über die Dichte der Logomere auf den jeweiligen Flächen deren Adhäsionsstärke eingestellt werden, zum anderen kann über das AT/GC Verhältnis der Logomere deren Schmelzpunkt variiert werden, so daß verklebte Flächen bei unterschiedlichen Temperaturen ihre Adhäsion verlieren. Aufgrund der Tatsache, daß hier unschädliche, leicht abbaubare Biomoleküle als Klebstoff verwendet werden, können derartige Klebstoffe auch zur medizinischen Verwendung (z. B. Chirurgie, Mikrochirurgie) eingesetzt werden. Eine andere Verwendung ist die Nutzung des beschriebenen Verfahrens in Hochpräzisions- und Authentifizierungsanwendungen.

Versetzt man die zur Herstellung von Logomeren verwendeten Algomeren mit Sequenzen zur Bindung von Liganden, so können stärkere Adhäsionswirkungen erreicht werden. Solche Liganden können im Falle von DNA-DNA-bindenden Proteinen, z. B. spezifische, gegen DNA-Sequenzen gerichtete Antikörper sein. Diese können z. B. jeweils über einen weiteren Liganden untereinander verbunden werden (siehe Abb. 19). Eine einfache, nichtprogrammierbare Adhäsionswirkung kann auch ohne Logomere erreicht werden, wenn die verwendeten Proteine (z. B. Antikörper) ihrerseits an die zu verklebenden Flächen binden können (wie z. B. Antikörper, wenn die zu verklebende Fläche ihr Antigen ist). Siehe dazu auch Abb. 20. Einfachere Klebstoffe sind möglich, wenn, an bestimmte Flächen spezifisch bindende, Proteine gentechnisch hergestellt werden.

Mit Logomeren können Oberflächen unterschiedlichster Struktur und unterschiedlichster physikalischer Eigenschaften hergestellt werden. Den Anwendungen ist gemein, daß sich mit den Logomeren nahezu beliebige Muster im Nanometerbereich bilden lassen, die das Skelett kleinster molekularer Bauelemente sein können.

Ein weiterer Gegenstand der vorliegenden Erfindung ist daher die Verwendung informationstragender Polymere, insbesondere der erfindungsgemäß erhältlichen informationstragenden Polymere, zur Herstellung oder Bearbeitung kleinsten molekularen Strukturen oder als molekularer Kleber.

# DE 199 14 808 A 1

## Referenzen

### Patente

5 International Publication Number	Inventor	Title
US4683202	Kary B. Mullis, Kensington, Calif.	Process for amplifying nucleic acid sequences
10 WO 97/07440	Adleman, Leonard, M; 18262 Hastings Way, Northridge, CA 91326 (US)	MOLECULAR COMPUTER
15 WO 97/29117	GUARNIERI, Frank [US/US]; 62 Lake Street, Brooklyn, NY 11223 (US). BANCROFT, Frank, Carter [US/US]; 51 Dewey Street, Huntington Station, NY 11743 (US).	A DNA-BASED COMPUTER
20 US5804373	Schweitzer; Allan Lee, Plainsboro, NJ Smith; Warren D., Plainsboro, NJ	Molecular automata utilizing single- or double-strand oligonucleotides
25		

30

### Veröffentlichungen

- L. M. Adleman, Molecular Computation of Solutions to Combinatorial Problems, *Science*, 266, 1021–1024, (1994)  
 Breslauer, K. J., Frank, R., Blocker, H., Marky, L. A., *Proc. Natl. Acad. Sci.*, 83, 3746–3750, 1989  
 Capecchi M. R., Altering the genome by homologous recombination. *Science*, 244(4910), 1288–1292, (1989)  
 35 Chomsky, N., Three models for the description of language, *JACM*, 2 : 3, 113–124, (1956)  
 Chomsky, N., On certain formal properties of grammars, *Inf. and Control*, 2 : 2, 137–167, (1959)  
 Chomsky, N., Formal properties of grammars, *Handbook of Math. Psych.*, 2, 323–418, (1963)  
 Frank Guarnieri, Makiko Fliss, Carter Bancroft, Making DNA Add, *Science*, 273, 220–223, (1996)  
 John E. Hopcroft, *Formal Languages And Their Relation To Automata*, (1969)  
 40 Jeffreys, Minisatellite repeat coding as a digital approach to DNA typing, *Nature*, 354, 204–209, (1991)  
 Kilby, N. J., Snaith, M. R. & Murray, J. A., Site-specific recombinases: tools for genome engineering, *Trends Genet.*, 9, 413–421, (1993)  
 Rolf Knippers, *Molekulare Genetik*, Georg Thieme Verlag, (1997)  
 Benjamin Lewin, *Genes V*, Oxford University Press, (1994)  
 45 Lund, A. H., Duch, M., Pedersen, F. S. Increased cloning efficiency by temperature-cycle ligation, *Nucleic Acids Research*, 24: (4), 800–801, (1996)  
 J. Sambrook, E. F. Fritsch, T. Maniatis, *Molecular Cloning, A Laboratory Manual*, (1989)  
 Jens Niehaus, DNA Computing: Bewertung und Simulation, Diplomarbeit am Fachbereich Informatik der Universität Dortmund, Lehrstuhl XI, 116–123, (1998)  
 50 Qi Ouyang, Peter D. Kaplan, Shumao Liu, Albert Libchaber, DNA Solution of the Maximal Clique Problem, *Science*, 278, 446–449, (1997)  
 Erik Winfree, Xiaoping Yang, Nadrian C. Seeman, Universal Computation via Self-assembly of DNA: Some Theory and Experiments, Proceedings of the 2nd DIMACS Meeting on DNA Based Computers, Princeton University, June 20–12, (1996)  
 55 Erik Winfree, Furong Liu, Lisa A. Wenzler & Nadrian C. Seeman, Design and self-assembly of two-dimensional DNA crystals, *Nature*, 394, 539–544, (1998)

### Patentansprüche

- 60 1. Verfahren zur Herstellung von informationstragenden Polymeren, das umfaßt,  
 I. eine reguläre Grammatik  $G = (\Sigma, V, R, S)$  mit einem endlichen Terminalalphabet  $\Sigma$ , einer endlichen Variablenmenge  $V$ , einer endlichen Regelmenge  $R$  und einem Startsymbol  $S$  zu definieren;  
 II. das NFR-Verfahren (Niehaus-Feldkamp-Rauhe-Verfahren) zur Herstellung von Monomersequenzen;  
 III. mit dem NFR-Verfahren eine in Schritt I definierte Grammatik zu implementieren, indem damit Monomersequenzen hergestellt werden, die die Regelmenge  $R$  einer Grammatik  $G$  eindeutig darstellen;  
 IV. aus den in Schritt III hergestellten Monomersequenzen für jede Regel der Regelmenge  $R$  von  $G$  ein die Regel repräsentierendes Oligomer zusammenzusetzen;  
 V. die in Schritt IV zusammengesetzten Oligomere zu informationstragenden Polymeren zu verknüpfen.

# DE 199 14 808 A 1

2. Verfahren nach Anspruch 1, dadurch gekennzeichnet, daß das Terminalalphabet  $\Sigma$  einer Grammatik G die Terminate 0, 1, n Startterminale s ( $s_0, s_1, \dots, s_n$ ) und m Endterminale e ( $e_0, e_1, \dots, e_m$ ), enthält, wobei n und m jeweils größer oder gleich 0 sind und eine natürliche Zahl darstellen. 5
3. Verfahren nach Anspruch 1 oder 2, dadurch gekennzeichnet, daß die in Schritt III konstruierte Monomersequenz Nukleotide, insbesondere Ribonukleotide, besonders bevorzugt Desoxyribonukleotide umfaßt.
4. Verfahren nach Anspruch 3, dadurch gekennzeichnet, daß die in Schritt III konstruierte Monomersequenz Proteinerkennungssequenzen (z. B. Restriktionsschnittstellen, Proteinbindungsstellen, Stopcodons) umfaßt. 10
5. Verfahren nach einem der vorhergehenden Ansprüche, dadurch gekennzeichnet, daß man die Synthese der Monomersequenzen in Schritt II und III in vitro, vorzugsweise mittels eines Oligonukleotid-Synthetisers vornimmt.
6. Verfahren zur Isolierung und Vervielfältigung von informationstragenden Polymeren, die nach einem der vorhergehenden Ansprüche erhalten wurden, dadurch gekennzeichnet, daß man in Schritt V erhaltene informationstragende Polymere in Klonierungsvektoren ligiert, kompetente Zellen mit diesen Vektoren transformiert und die erfolgreich transformierten Bakterien anhand von Selektionsmarkern selektioniert. 15
7. Verfahren zum Auslesen von Information aus informationstragenden Polymeren, die nach einem der vorhergehenden Ansprüche erhalten oder isoliert und vervielfältigt wurden, dadurch gekennzeichnet, daß man
- a) mindestens n-1 das informationstragende Polymer enthaltende Lösungen mit jeweils einem Paar gegenüberliegender Primer versetzt, wobei n für die Zahl der als Elongatoren in dem Polymer enthaltenen Oligomere steht;
  - b) mindestens n-1 PCR-Ansätze durchführt, wobei n für die Zahl der als Elongatoren in dem Polymer enthaltenen Oligomere steht und jeweils ein Primer eines jeden Paares in dem dem Elongator gegenüberliegenden Terminator primt und der andere Primer in dem Elongator selbst primt; 20
  - c) die aus der PCR erhaltenen Polymer-Fragmente nach ihrer Länge durch Elektrophorese auf trennt und
  - d) das aus der Elektrophorese erhaltene Muster optisch ausliest.
8. Verfahren nach Anspruch 7, dadurch gekennzeichnet, daß das Auslesen in Schritt d) automatisiert durch einen Scanner oder eine Sequenziemaschine erfolgt. 25
9. Informationstragendes Polymer, erhältlich nach einem der Ansprüche 1 bis 6.
10. Zufallszahlengenerator, umfassend informationstragende Polymere, insbesondere Polymere nach Anspruch 9.
11. Polymerer Datenspeicher, umfassend informationstragende Polymere nach Anspruch 9.
12. DNA-Computer, umfassend informationstragende Polymere nach Anspruch 9.
13. Verwendung von informationstragenden Polymeren nach Anspruch 9 zur Herstellung von Molekulargewichtsstandards. 30
14. Verwendung von informationstragenden Polymeren nach Anspruch 9 zur Darstellung von Datenstrukturen.
15. Verwendung von informationstragenden Polymeren, insbesondere von Polymeren nach Anspruch 9 als Marker oder Signaturen.
16. Verwendung von informationstragenden Polymeren, insbesondere von Polymeren nach Anspruch 9 zum Zwecke der Qualitätssicherung. 35
17. Verwendung von informationstragenden Polymeren, insbesondere von Polymeren nach Anspruch 9 zum Zwecke der Fälschungssicherung.
18. Verwendung von informationstragenden Polymeren, insbesondere von Polymeren nach Anspruch 9 zum Zwecke der Kennzeichnung gentechnischer Erzeugnisse.
19. Verwendung von informationstragenden Polymeren, insbesondere von Polymeren nach Anspruch 9 zum Zwecke der Kennzeichnung von Lebensmitteln. 40
20. Verwendung von informationstragenden Polymeren, insbesondere von Polymeren nach Anspruch 9 zum Zwecke der Kennzeichnung von Organismen.
21. Verwendung von informationstragenden Polymeren, insbesondere von Polymeren nach Anspruch 9 zum Zwecke der Kennzeichnung chemischer Erzeugnisse.
22. Verwendung von informationstragenden Polymeren, insbesondere von Polymeren nach Anspruch 9 zum Zwecke der Kennzeichnung medizinischer und pharmazeutischer Erzeugnisse. 45
23. Verwendung von informationstragenden Polymeren, insbesondere von Polymeren nach Anspruch 9 zum Zwecke der Kennzeichnung von Dokumenten.
24. Verwendung von informationstragenden Polymeren, insbesondere von Polymeren nach Anspruch 9 zum Zwecke der Kennzeichnung von Geld. 50
25. Verwendung von informationstragenden Polymeren, insbesondere von Polymeren nach Anspruch 9 zum Zwecke der Kennzeichnung von Gegenständen und Maschinen.
26. Verwendung von informationstragenden Polymeren, insbesondere von Polymeren nach Anspruch 9 zum Zwecke der Kennzeichnung von Flüssigkeiten, Lösungen, Suspensionen oder Emulsionen. 55
27. Verwendung von informationstragenden Polymeren, insbesondere von Polymeren nach Anspruch 9 zur Verschlüsselung von Information.
28. Verwendung von informationstragenden Polymeren, insbesondere von Polymeren nach Anspruch 9 zum Zwecke der Authentisierung von Personen und Gegenständen.
29. Verwendung von informationstragenden Polymeren, insbesondere von Polymeren nach Anspruch 9 als molekulare Kleber. 60
30. Verwendung von informationstragenden Polymeren nach Anspruch 9 zur Herstellung oder Bearbeitung kleinsten molekularen Strukturen.
31. Verwendung von informationstragenden Polymeren nach Anspruch 9 zur Qualitätskontrolle synthetisch hergestellter Oligonukleotide. 65

Hierzu 12 Seite(n) Zeichnungen

**- Leerseite -**

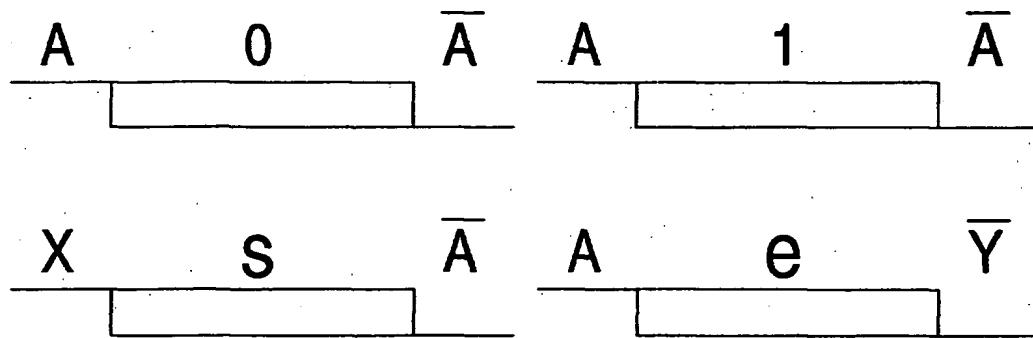


Abbildung 1

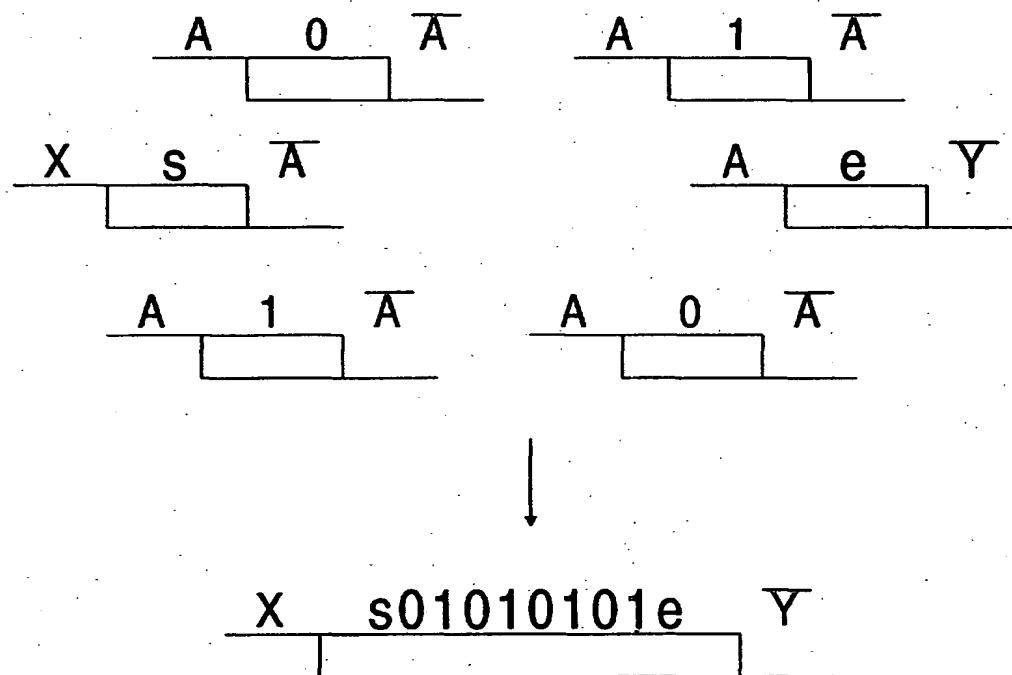


Abbildung 2

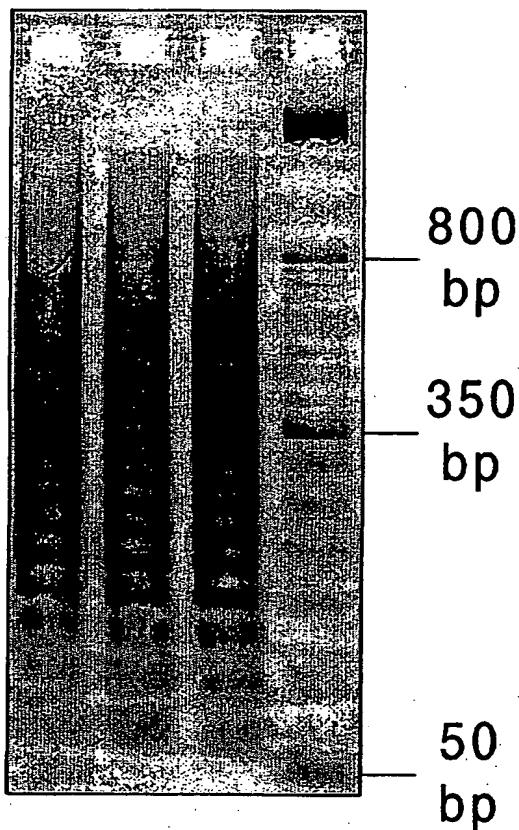


Abbildung 3

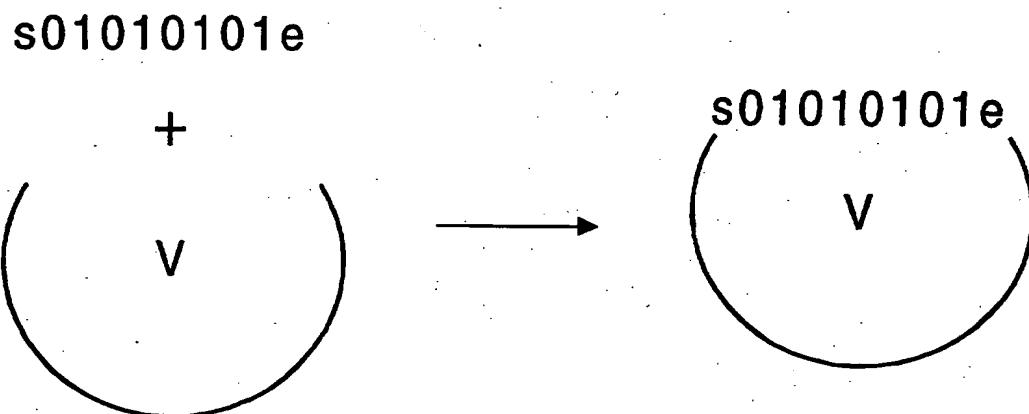


Abbildung 4

→  
s01010101e s01010101e

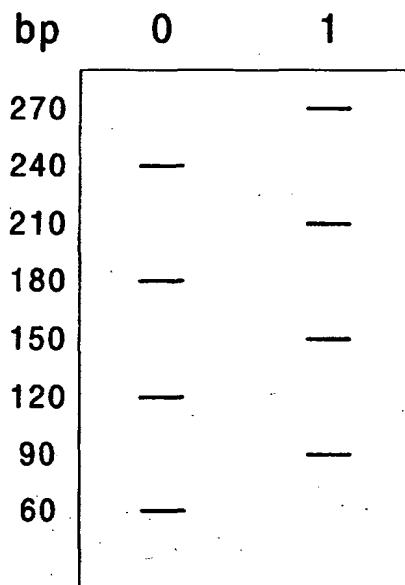


Abbildung 5

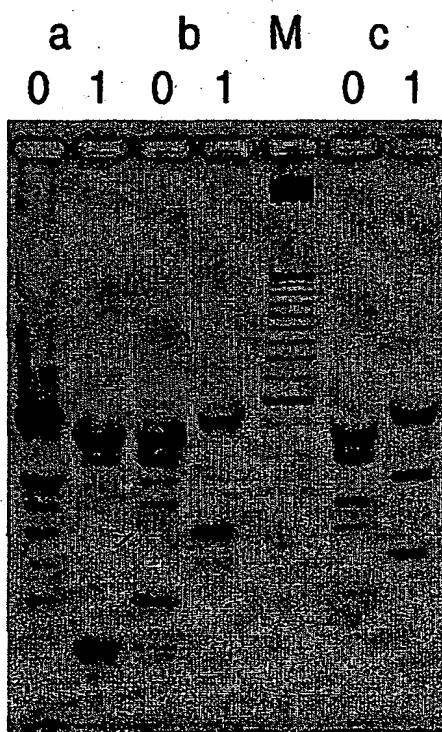


Abbildung 6

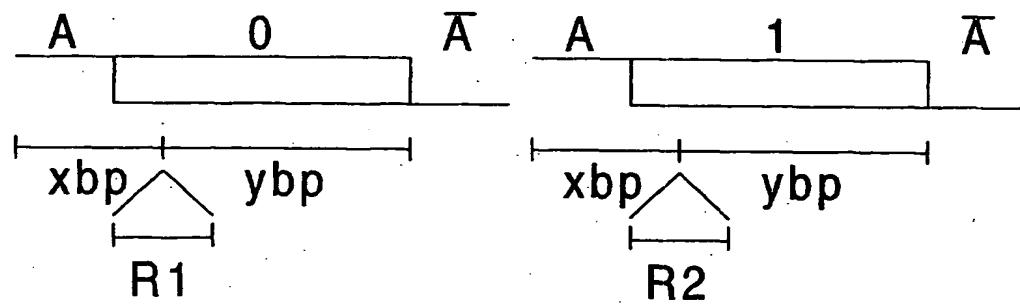


Abbildung 7

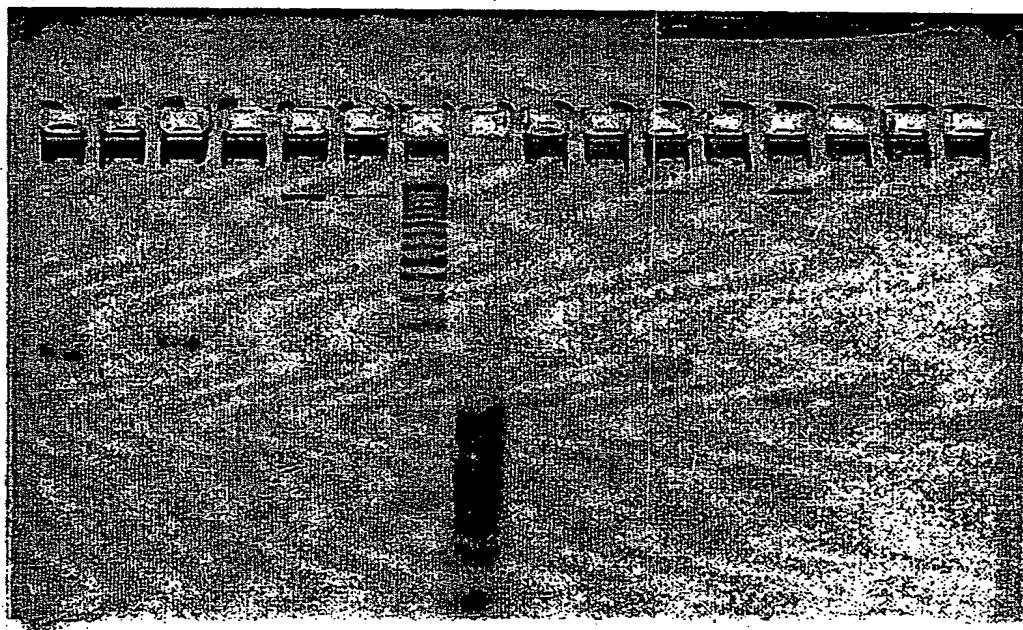


Abbildung 8

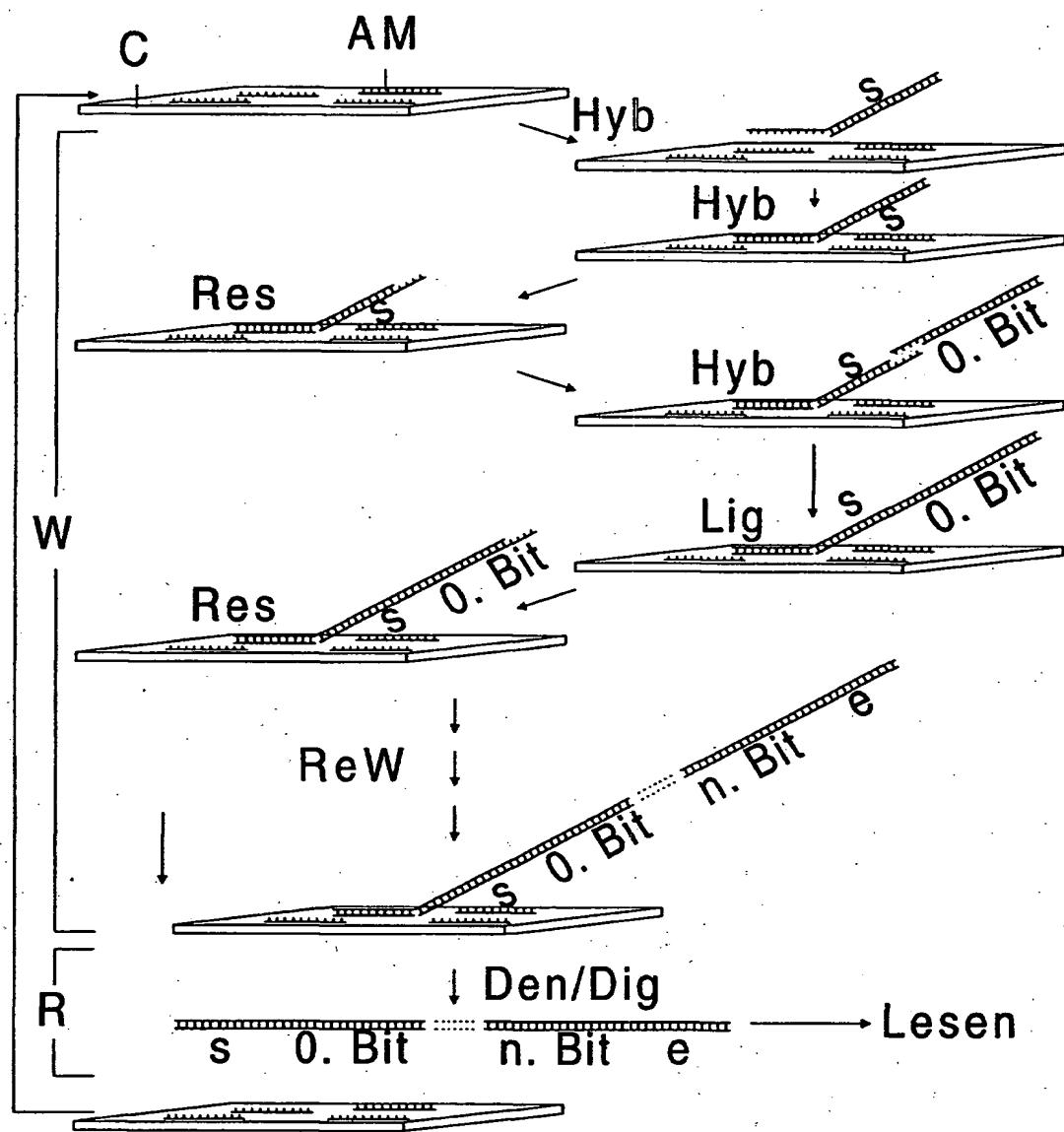


Abbildung 9

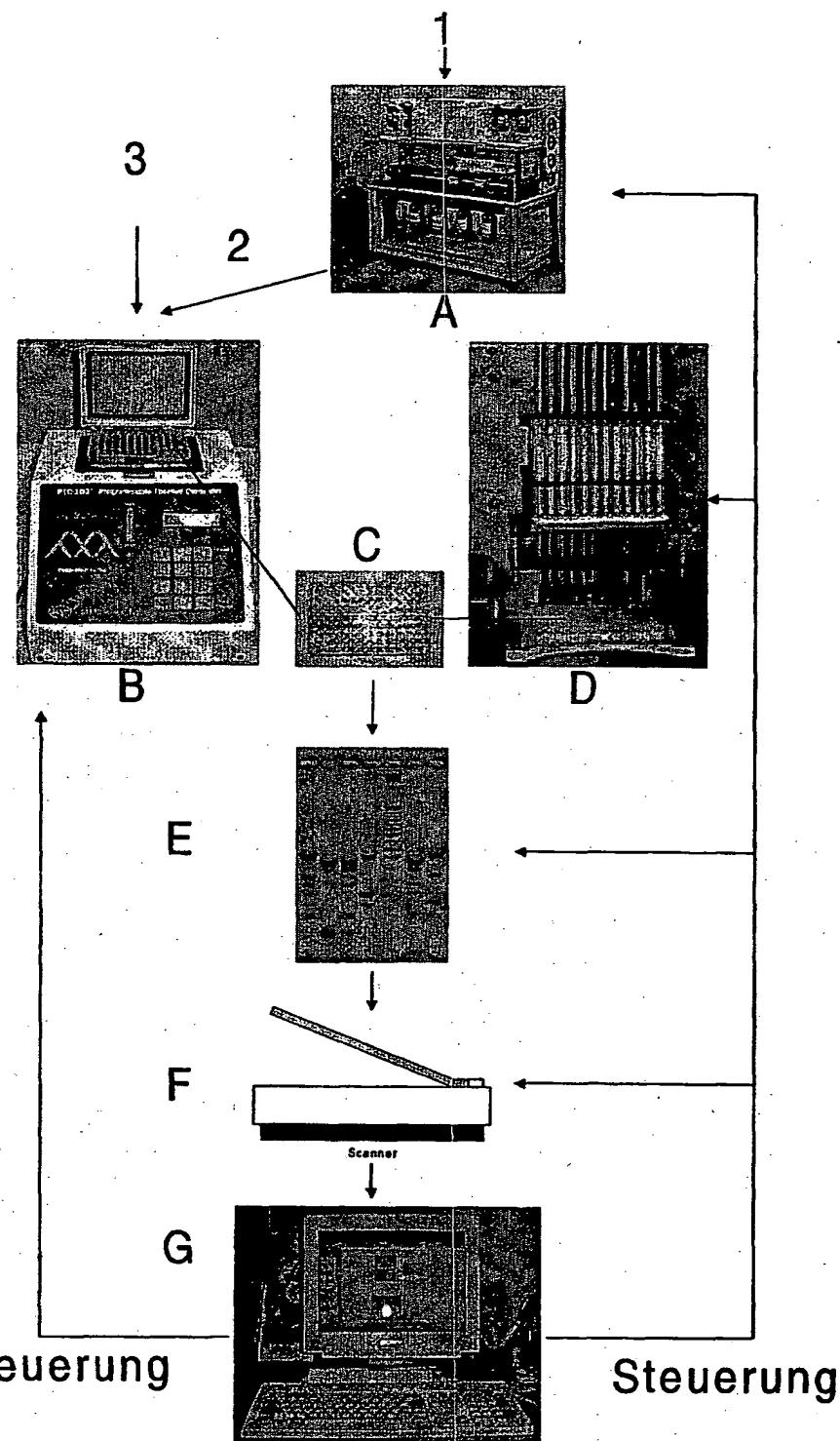


Abbildung 10

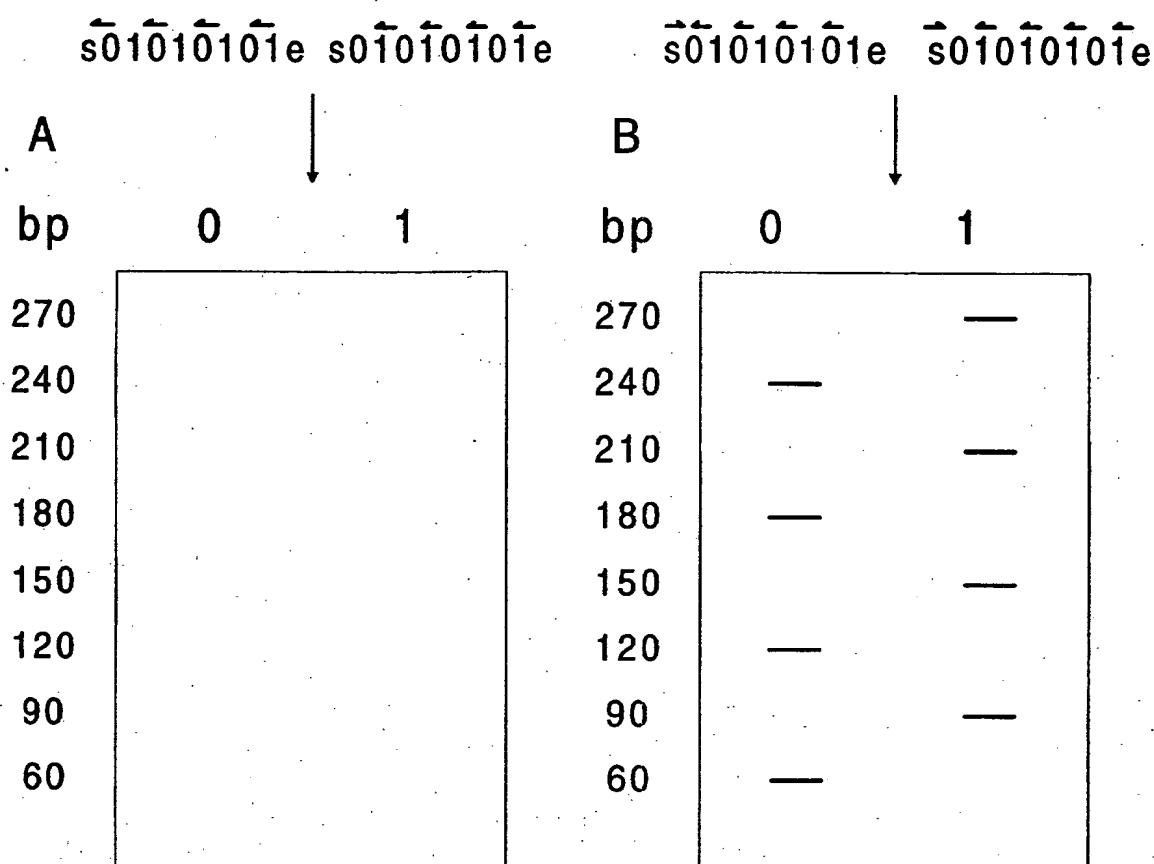


Abbildung 11

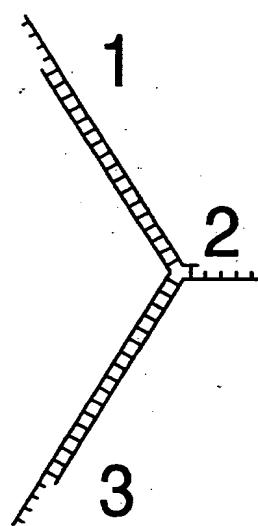


Abbildung 12

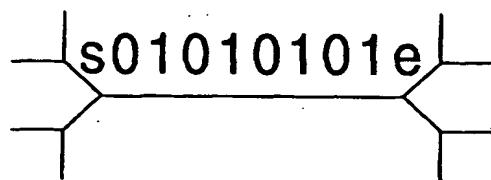


Abbildung 13

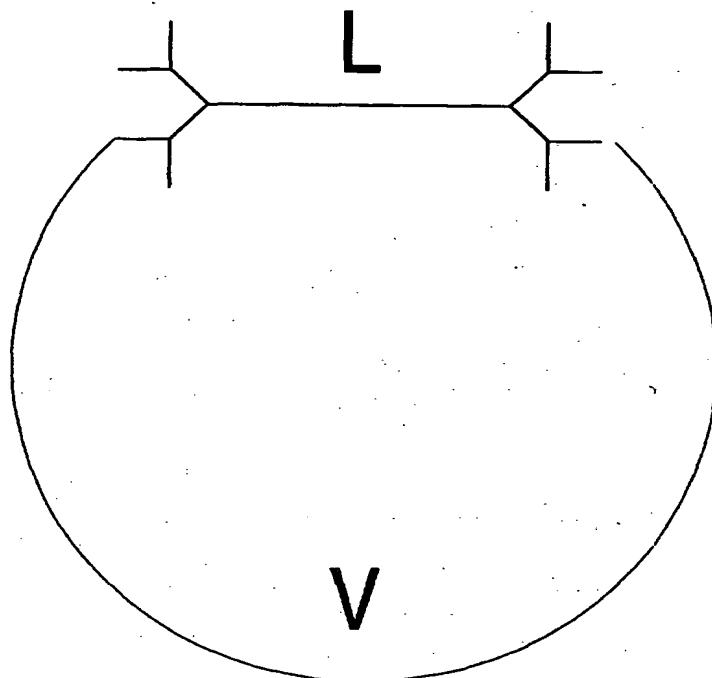


Abbildung 14

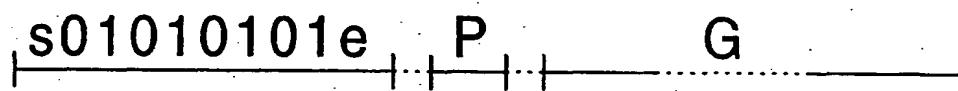
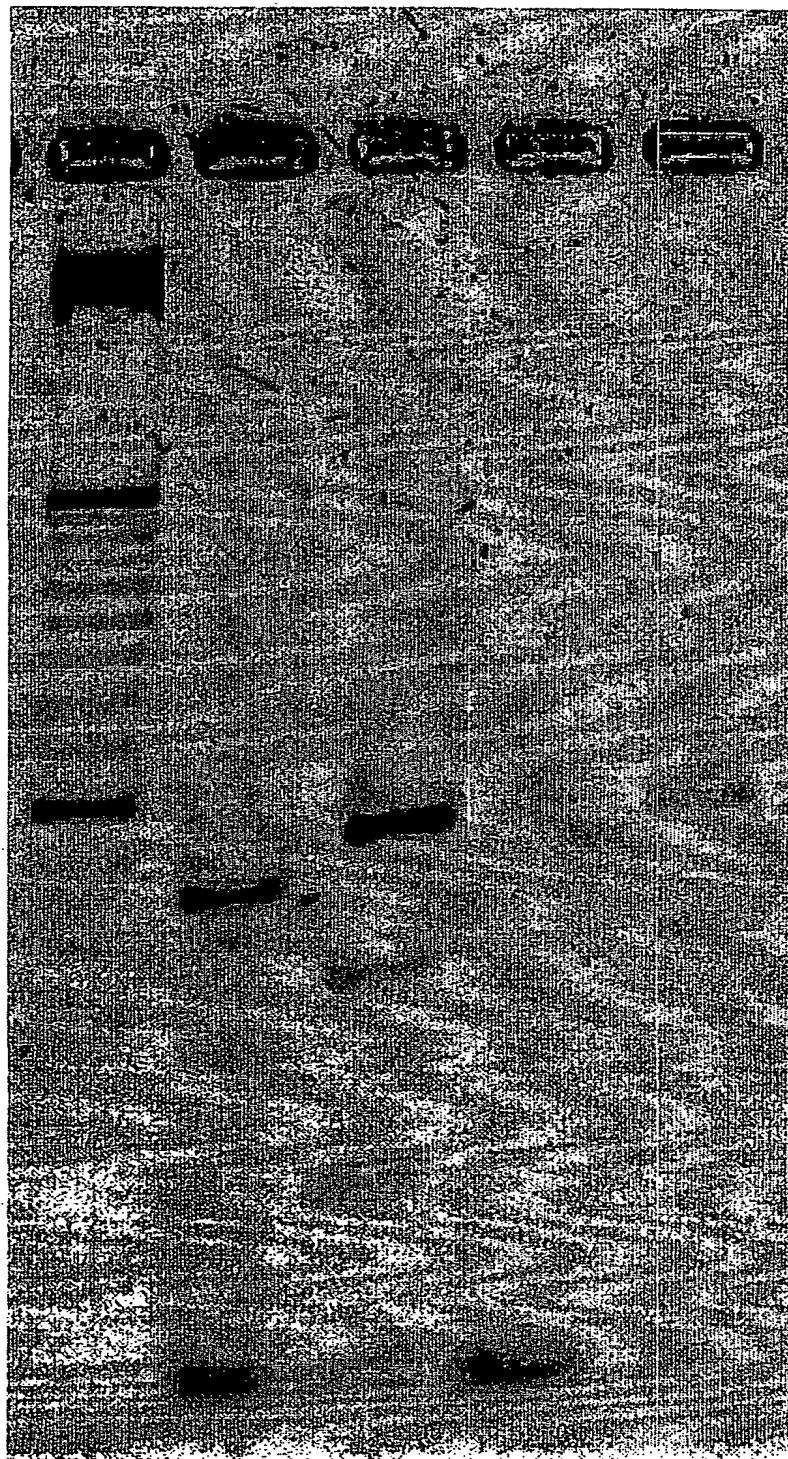


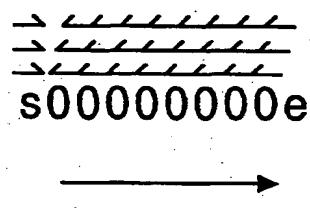
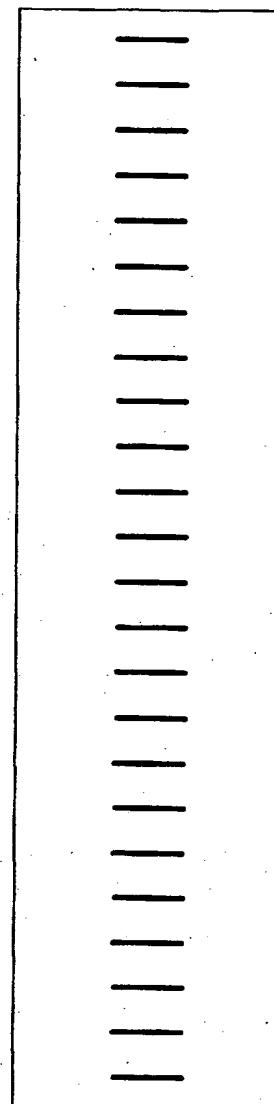
Abbildung 15



**Abbildung 16**

2 bp 0

290  
280  
270  
260  
250  
240  
230  
220  
210  
200  
190  
180  
170  
160  
150  
140  
130  
120  
110  
100  
90  
80  
70  
60



1

Abbildung 17

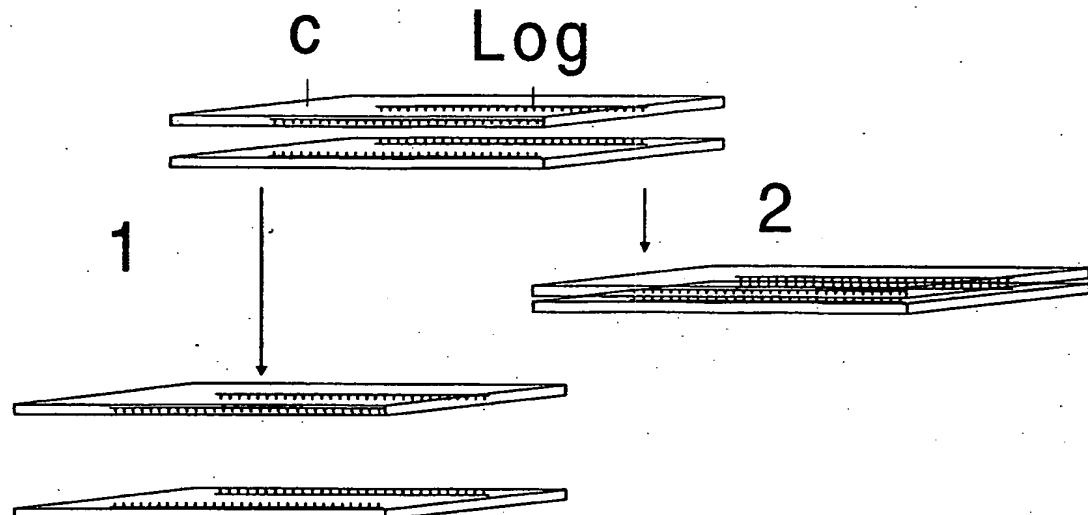


Abbildung 18

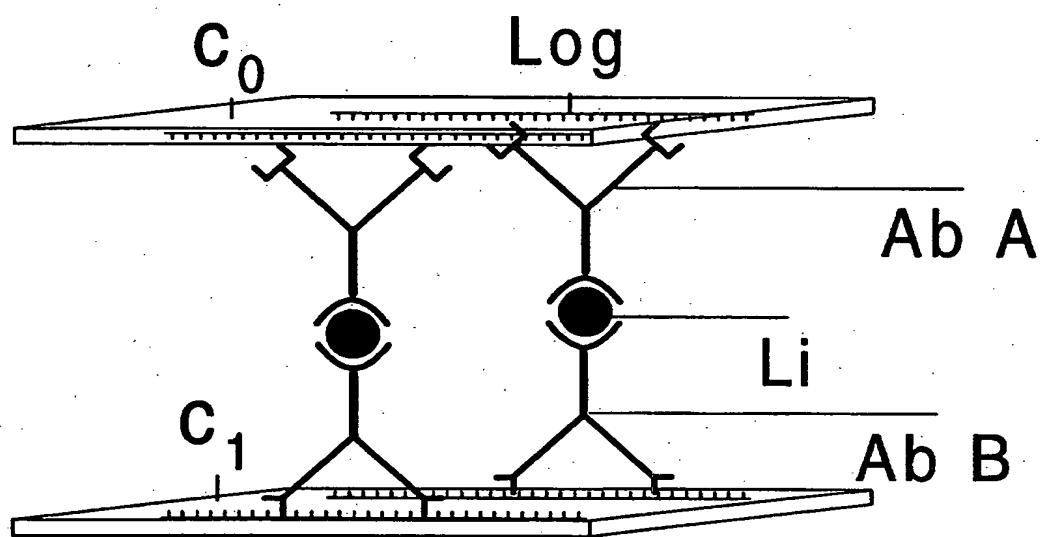


Abbildung 19

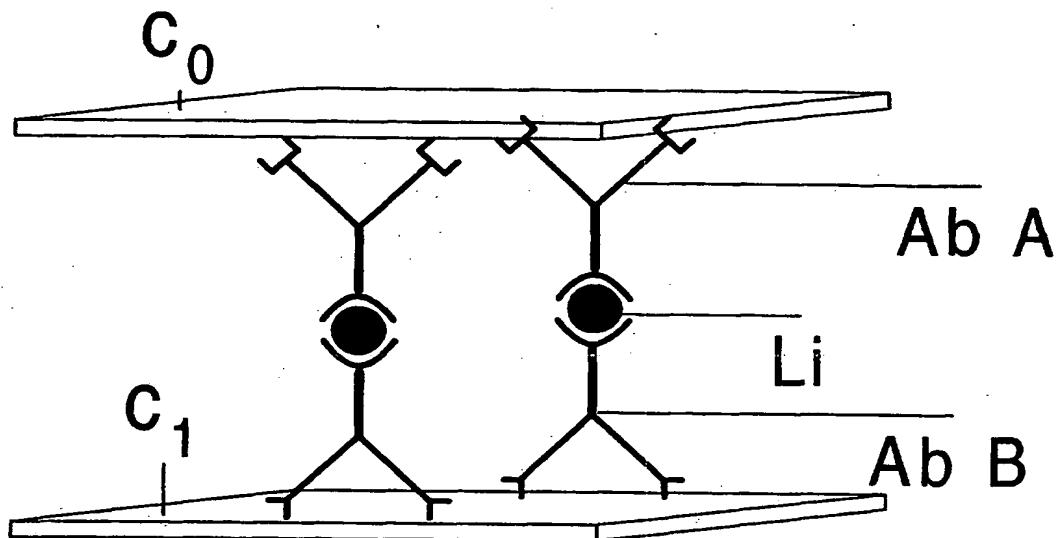


Abbildung 20